Physica A 429 (2015) 252–260

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

On the super-additivity and estimation biases of quantile contributions

Nassim Nicholas Taleb [a,*], Raphael Douady [b]

[a] School of Engineering, New York University, United States

[b] Riskdata & C.N.R.S. Paris, Labex ReFi, Centre d'Economie de la Sorbonne, France

HIGHLIGHTS

- Estimating concentration (inequality or dispersion) or other statistical properties (such as severity of violent conflicts) from top quantile contributions is inconsistent under aggregation.
- The measure increases with the size of the total population and converges very slowly.
- The bias is more acute at fatter tails, lower tail exponent alpha and smaller centile.
- The weighted average of measures for A and B will be $\leq$ than that from $A \cup B$.
- The effect is exacerbated under mixing distributions (stochastic tail exponent).

ARTICLE INFO

Article history:
Received 12 May 2014
Received in revised form 14 November 2014
Available online 10 February 2015

Keywords:
Quantile contribution
Measure of inequality
Estimation bias
Power law
Law of large number
Concentration measure

ABSTRACT

Sample measures of top centile contributions to the total (concentration) are downward biased, unstable estimators, extremely sensitive to both sample and population size and concave in accounting for large deviations. It makes them particularly unfit in domains with power law tails, especially for low values of the exponent. These estimators can vary over time and increase with the population size, thus providing the illusion of structural changes in concentration. They are also inconsistent under aggregation and mixing distributions, as the weighted average of concentration measures for A and B will tend to be lower than that from $A \cup B$. In addition, it can be shown that under such fat tails, increases in the total sum need to be accompanied by increased sample size of the concentration measurement. We examine the estimation superadditivity and bias under homogeneous and mixed distributions.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Vilfredo Pareto noticed that 80% of the land in Italy belonged to 20% of the population, and vice-versa, thus both giving birth to the power law class of distributions and the popular saying 80/20. The self-similarity at the core of the property of power laws [1,2] allows us to recurse and reapply the 80/20 to the remaining 20%, and so forth until one obtains the result that the top percent of the population will own about 53% of the total wealth.

It looks like such a measure of concentration can be seriously biased, depending on how it is measured, so it is very likely that the true ratio of concentration of what Pareto observed, that is, the share of the top percentile, was closer to 70%, hence changes year-on-year would drift higher to converge to such a level from larger sample. In fact, as we will show in this discussion, for, say wealth, more complete samples resulting from technological progress, and/or larger population and

* Corresponding author.
E-mail address: NNT1@NYU.edu (N.N. Taleb).

http://dx.doi.org/10.1016/j.physa.2015.02.038
0378-4371/© 2015 Elsevier B.V. All rights reserved.

economic growth will make such a measure converge by increasing over time, for no other reason than expansion in sample space or aggregate value.

The core of the problem is that, for the class one-tailed fat-tailed random variables, that is, bounded on the left and unbounded on the right, where the random variable $X \in [x_{\min}, \infty)$, the in-sample quantile contribution is a biased estimator of the true value of the actual quantile contribution.

Let us define the *quantile contribution*

$$\kappa_q = q \frac{\mathbb{E}[X|X > h(q)]}{\mathbb{E}[X]}$$

where $h(q) = \inf\{h \in [x_{min}, +\infty)\, , \mathbb{P}(X > h) \leq q\}$ is the exceedance threshold for the probability q.

For a given sample $(X_k)_{1 \leq k \leq n}$, its "natural" estimator $\widehat{\kappa_q} \equiv \frac{qth\ \text{percentile}}{total}$, used in most academic studies, can be expressed, as

$$\widehat{\kappa_q} \equiv \frac{\sum_{i=1}^{n} \mathbb{1}_{X_i > \hat{h}(q)} X_i}{\sum_{i=1}^{n} X_i}$$

where $\hat{h}(q)$ is the estimated exceedance threshold for the probability q:

$$\hat{h}(q) = \inf \left\{ h : \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x>h} \leq q \right\}.$$

We shall see that the observed variable $\widehat{\kappa_q}$ is a downward biased estimator of the true ratio κ_q, the one that would hold out of sample, and such bias is in proportion to the fatness of tails and, for very fat tailed distributions, remains significant, even for very large samples.

2. Estimation for unmixed Pareto-tailed distributions

Let X be a random variable belonging to the class of distributions with a "power law" right tail, that is:

$$\mathbb{P}(X > x) \sim L(x)\, x^{-\alpha} \tag{1}$$

where $L : [x_{\min}, +\infty) \to (0, +\infty)$ is a slowly varying function, defined as $\lim_{x \to +\infty} \frac{L(kx)}{L(x)} = 1$ for any $k > 0$.

There is little difference for small exceedance quantiles (<50%) between the various possible distributions such as Student's t, Lévy α-stable, Dagum, [3,4] Singh–Maddala distribution [5], or straight Pareto.

For exponents $1 \leq \alpha \leq 2$, as observed in Ref. [6], the law of large numbers operates, though *extremely* slowly. The problem is acute for α around, but strictly above 1 and severe, as it diverges, for $\alpha = 1$.

2.1. Bias and convergence

2.1.1. Simple Pareto distribution

Let us first consider $\phi_\alpha(x)$ the density of a α-Pareto distribution bounded from below by $x_{\min} > 0$, in other words: $\phi_\alpha(x) = \alpha x_{\min}^{\alpha} x^{-\alpha-1} \mathbb{1}_{x \geq x_{\min}}$, and $\mathbb{P}(X > x) = \left(\frac{x_{\min}}{x}\right)^{\alpha}$. Under these assumptions, the cutpoint of exceedance is $h(q) = x_{\min}\, q^{-1/\alpha}$ and we have:

$$\kappa_q = \frac{\int_{h(q)}^{\infty} x\, \phi(x)\, dx}{\int_{x_{min}}^{\infty} x\, \phi(x)\, dx} = \left(\frac{h(q)}{x_{\min}}\right)^{1-\alpha} = q^{\frac{\alpha-1}{\alpha}}. \tag{2}$$

If the distribution of X is α-Pareto only beyond a cut-point x_{cut}, which we assume to be below $h(q)$, so that we have $\mathbb{P}(X > x) = \left(\frac{\lambda}{x}\right)^{\alpha}$ for some $\lambda > 0$, then we still have $h(q) = \lambda q^{-1/\alpha}$ and

$$\kappa_q = \frac{\alpha}{\alpha - 1} \frac{\lambda}{\mathbb{E}[X]} q^{\frac{\alpha-1}{\alpha}}.$$

The estimation of κ_q hence requires that of the exponent α as well as that of the scaling parameter λ, or at least its ratio to the expectation of X.

Table 1 shows the bias of $\widehat{\kappa_q}$ as an estimator of κ_q in the case of an α-Pareto distribution for $\alpha = 1.1$, a value chosen to be compatible with practical economic measures, such as the wealth distribution in the world or in a particular country, including developed ones.[1] In such a case, the estimator is extremely sensitive to "*small*" samples, "*small*" meaning in

[1] This value, which is lower than the estimated exponents one can find in the literature – around 2 – is, following [7], a lower estimate which cannot be excluded from the observations.

Download English Version:

https://daneshyari.com/en/article/974439

Download Persian Version:

https://daneshyari.com/article/974439

Daneshyari.com