A SMOOTH TRANSFORMATION OF PSEUDO-POWER LAWS

A.1 ADJUSTING FOR BOUNDED SUPPORT

In many situations with power laws, we use the terminology "infinite" to describe the "very large". In cases where "infinite" corresponds to an unknown upper (or lower) bound, it would be certainly proper to use *infinite*. But there are many cases of known upper bound, for which the notion of infinite can be misapplied. The distribution is no longer a power law; we are just using power law as terminology.

So "infinite" is often used when it is not technically adequate a designation in many cases. Further, the standard mathematical methods and proofs we have that were derived for power laws do not map to closed intervals.

Consider wars and violence: the distribution cannot be unbounded since the maximum amount of fatalities cannot exceed the world population; the mean cannot be technically "infinite". The same applies to problems in hydrology, size of companies (they cannot exceed the total GDP), forest fires, insurance claims, etc.

> In a conversation with the late Benoit Mandelbrot, we discussed the difference between "infinite support" and "large but finite" (with obviously known upper bound) and the effect on infinite mean/variance. We both agreed that the expectation would be at the upper bound. It seemed natural.
>
> It turned out that we were wrong, at least about the mean.
>
> Higher moments, of course will be pulled towards the upper bound.

Using X as the r.v. for number of incidences in war, consider a smooth rescaling function $\varphi : [L, H] \to [L, \infty)$ satisfying:

i φ is "smooth": $\varphi \in C^\infty$ [or φ is analytic (though not necessarily smooth), yet a stronger condition.],

ii $\varphi^{-1}(\infty) = H$,

iii $\varphi^{-1}(L) = L$,

which gives us:

$$\varphi(x) = L - H \log\left(\frac{H-x}{H-L}\right) \tag{A.1}$$

We can perform appropriate analytics on $x' = \varphi(x)$ given that it is unbounded, and properly fit power law exponents. Then we can rescale back for the properties of X.

The distribution of x can be rederived as follows from the distribution of X':

$$\int_L^\infty f(x')\, \mathrm{d}x' = \int_L^{\varphi^{-1}(\infty)} g(x)\, \mathrm{d}x, \tag{A.2}$$

where $\varphi^{-1}(u) = (L-H)e^{\frac{L-u}{H}} + H$.

We were surprised to discover that:

- "infinite mean" is not at the boundary H as we expected.

- one can actually get the kurtosis to see how unreliable the standard deviation. But mean deviation remains –sort of – reliable.

Let us see how the moments behave when they are "infinite" with the simplest example.

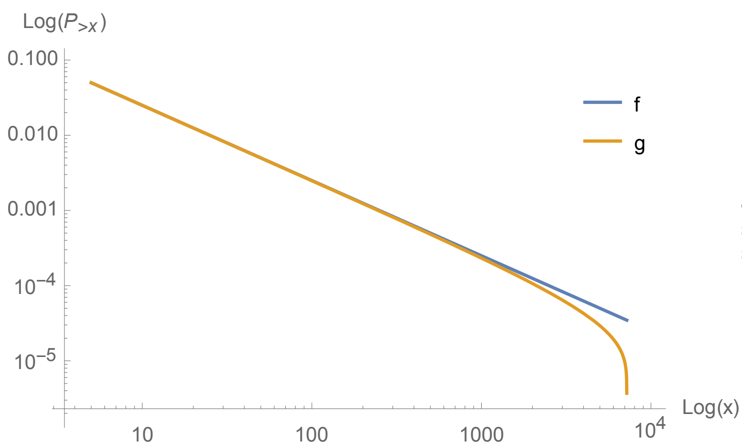

Figure A.1: Loglogplot comparison of f and g, showing smooth-pasting style boundary around H.

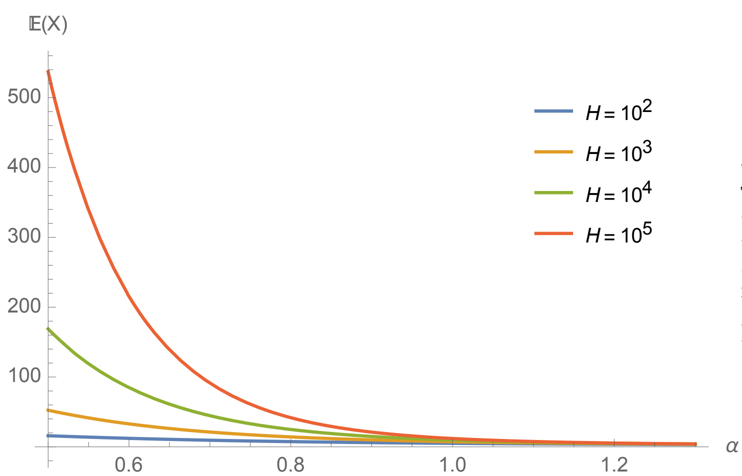

Figure A.2: We have $L = 1, H = 10^2, 10^3, 10^4, 10^5$. The mean doesn't seem to be "infinite" or close to the boundary, even with $\alpha < 1$. We get the maximal values of 15.91, 52.4, 169, 537, respectively, with $\alpha = \frac{1}{2}$.

A.2 SIMPLEST CASE OF STRAIGHT PARETO

(Note that we are using the lowercase f both for the PDF and the distribution.) Where $x' = \varphi(x)$. From

$$f(x') = \alpha L^{\alpha} x^{-\alpha-1}, \; x' \in [L, \infty) \tag{A.3}$$

we get the transformed

$$g(x) = \frac{\alpha H L^{\alpha} \left(H \log\left(\frac{H-L}{H-x}\right) + L \right)^{-\alpha-1}}{H-x}, \; x \in [L, H] \tag{A.4}$$

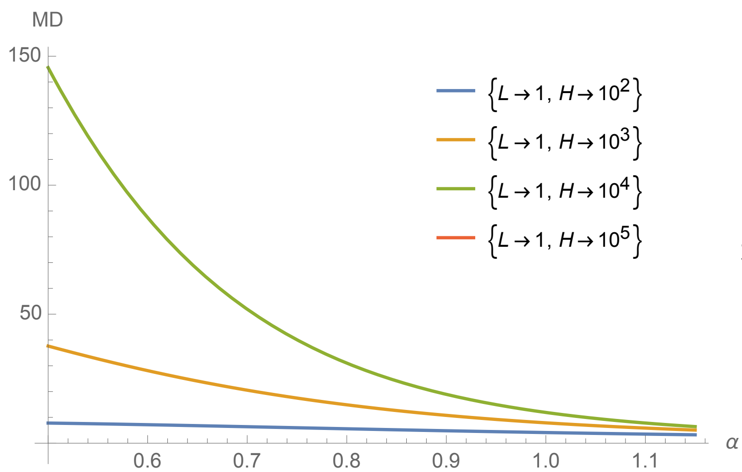

Figure A.3: Mean Deviation for $L = 1, H = 10^3, 10^4, 10^5$.

with first moment

$$\mathbb{E}(X) = H - \alpha e^{L/H}(H-L)E_{\alpha+1}\left(\frac{L}{H}\right) \tag{A.5}$$

where E_.(.) is the exponential integral $E_n(z) = \int_1^\infty \frac{e^{-zt}}{t^n} dt$ also $E_n(z) = x^{n-1}\Gamma(1-n,z)$.

Second moment

$$\mathbb{E}(X^2) = \alpha e^{L/H}(H-L)\left(e^{L/H}(H-L)E_{\alpha+1}\left(\frac{2L}{H}\right) - 2HE_{\alpha+1}\left(\frac{L}{H}\right)\right) + H^2 \tag{A.6}$$

Fourth moment

$$\begin{aligned}\mathbb{E}(X^4) = \alpha e^{L/H}(H-L)\Bigg(e^{L/H}(H-L)\left(6H^2E_{\alpha+1}\left(\frac{2L}{H}\right) + e^{L/H}(H-L)\right.\\ \left.\left(e^{L/H}(H-L)E_{\alpha+1}\left(\frac{4L}{H}\right) - 4HE_{\alpha+1}\left(\frac{3L}{H}\right)\right)\right) - 4H^3E_{\alpha+1}\left(\frac{L}{H}\right)\Bigg) + H^4\end{aligned} \tag{A.7}$$

$$\mathbb{P}(X < \mathbb{E}(X)) = 1 - L^\alpha\left(H\left(\alpha\log\left(\frac{H}{L}\right) - \log\left(\alpha\Gamma\left(-\alpha, \frac{L}{H}\right)\right)\right)\right)^{-\alpha} \tag{A.8}$$

A.3 MORE GENERAL CASE OF GENERALIZED BETA 2^{nd} KIND

Note that L needs to match the left support of the distribution so, in this case, L is necessarily 0. Should we need a lower bound at L, would have to rescale by changing the random variable to $x + L$ or, better, $(x+L)\frac{H-L}{H}$.

$$f(x') = \frac{\alpha\left(\frac{x'}{\beta}\right)^{\alpha p-1}\left(\left(\frac{x}{\beta}\right)^\alpha + 1\right)^{-p-q}}{\beta B(p,q)},\ x' > 0$$

We can parametrize p, q, and β to get the appropriate subdistribution. For instance it simplifies to the Dagum distribution$[p, \alpha, \beta]$ when $q = 1$, to the Singh Maddala distribution$[q, \alpha, \beta]$ when $p = 1$, and to the log logistic distribution$[\alpha, \beta]$ when both $p = 1$ and $q = 1$.

$$g(x) = \frac{\alpha H}{\beta \,(H-x)\, B(p,q)} \frac{\left(-\frac{H \log\left(1-\frac{x}{H}\right)}{\beta}\right)^{\alpha p-1}}{\left(\left(-\frac{H \log\left(1-\frac{x'}{H}\right)}{\beta}\right)^{\alpha}+1\right)^{p+q}}\;,\; x \in [0, H) \tag{A.9}$$

After a few manipulations

$$\mathbb{E}(X) = \frac{\alpha H}{\beta B(p,q)} \int_0^\infty \frac{\left(He^{-v}-H\right)\left(\frac{Hv}{\beta}\right)^{\alpha p-1}}{\left(\left(\frac{Hv}{\beta}\right)^{\alpha}+1\right)^{p+q}}\, dv \tag{A.10}$$

We cannot get explicit solutions without fixing a parameter, α or p and q. In the infinite mean case:

$$\mathbb{E}(X)|_{\alpha=1} = -\frac{H\,\Gamma(p)\left(U\left(p, 1-q, \frac{\beta}{H}\right)-\frac{\Gamma(q)}{\Gamma(p+q)}\right)}{B(p,q)}$$

where U is the confluent hypergeometric function $U(a,b,z) = \frac{1}{\Gamma(a)}\int_0^\infty t^{a-1}(t+1)^{-a+b-1}e^{t(-z)}\,\mathrm{d}t$. Infinite variance case:

$$\begin{aligned}\mathbb{E}(X)|_{\alpha=2} = & H\,{}_1F_2\left(p; \frac{1}{2}, 1-q; -\frac{\beta^2}{4H^2}\right) - \frac{1}{\Gamma(p)\Gamma(q)} \\ & - 2H\Gamma(-2q)\left(\frac{H}{\beta}\right)^{-2q}\Gamma(p+q)\left({}_1F_2\left(p+q; q+\frac{1}{2}, q+1; -\frac{\beta^2}{4H^2}\right)\right. \\ & \left. + \beta\Gamma\left(p+\frac{1}{2}\right)\Gamma\left(q-\frac{1}{2}\right)\,{}_1F_2\left(p+\frac{1}{2}; \frac{3}{2}, \frac{3}{2}-q; -\frac{\beta^2}{4H^2}\right) + H\Gamma(p)\Gamma(q)\right)\end{aligned} \tag{A.11}$$

A.3.1 Findings

i Kurtosis is too high to make STD reliable
ii MD appears also too high but reliable
iii Working on its distribution

A.4 CONCLUSION

Even with wildest powerlaws the mean stays well behaved, several order of magnitude below the hard upper bound.