Gintropy and the LGGR model for income and particle spectra

Entropy 22 (2020) 879

T.S. Biró[1,2] **Z. Néda[3]**

[1]MTA WIGNER Research Centre for Physics, Budapest, Hungary

[2]Complexity Science Hub, Vienna, Austria [3]Hungarian Institute for Physics, Babeş-Bolyai University Cluj, Romania

version: April 22, 2021

ACHT 2021 talk Apr. 21-23. 2021.

What is it?

What is gintropy

a textual definition

- Formulas resembling entropy terms
- by calculating the Gini-index *def. due to Corrado Gini*
- (Gross Inequality Natural Index)
- **Its properties are entropy-like, but it is not an entropy.**

Definitions

Gini index due to Corrado Gini

Let $P(x)$ be a normalized PDF. The GINI is defined as:

$$G \equiv \frac{\langle\, |x-y|\, \rangle}{\langle\, |x+y|\, \rangle} = \frac{1}{2\,\langle x \rangle} \int\limits_0^\infty dx \int\limits_0^\infty dy\; |x-y|\; P(x)\, P(y). \tag{1}$$

Cumulant population fraction (from the rich end): $\overline{C}(0) = 1.$

$$\overline{C}(x) \equiv \int\limits_x^\infty dy\, P(y). \tag{2}$$

Cumulant wealth per average (from the rich end): $\overline{F}(0) = 1.$

$$\overline{F}(x) \equiv \frac{1}{\langle x \rangle} \int\limits_x^\infty dy\, y\, P(y). \tag{3}$$

Pareto point

80/20: 80% of the wealth is posessed by 20% of the folk.

Total population: $N_{\rm tot} = \int\limits_0^\infty dx\, N(x)$, total wealth: $X_{\rm tot} = \int\limits_0^\infty dx\, x\, N(x)$.

By this $P(x) = N(x)/N_{\rm tot}$ and $\langle x \rangle = X_{\rm tot}/N_{\rm tot}$.

At any p-Pareto point: $\overline{C}(x_P) = p$; and $\overline{F}(x_P) = (1-p)$

$(1-p)X_{\rm tot}$ **is owned by** $pN_{\rm tot}$**.**

The original definition (1) can be expressed by using the cumulatives as

$$G = \int\limits_0^\infty dx\, P(x) \int\limits_x^\infty dy\, \frac{y-x}{\langle x \rangle}\, P(y) = \int\limits_0^\infty dx\, P(x) \left[\overline{F}(x) - \frac{x}{\langle x \rangle} \overline{C}(x) \right] \qquad (4)$$

Proof uses the $x \leftrightarrow y$ symmetry under the integral.

GINI

alternative expressions $\int_0^\infty dx\, P(x) \ldots = \int_0^1 d\overline{C} \ldots$ $\int_0^\infty dx\, \frac{x}{\langle x \rangle} P(x) \ldots = \int_0^1 d\overline{F} \ldots$

The cumulative of the cumulative:

$$\overline{h}(x) \equiv \int_x^\infty dy\, \overline{C}(y) = \int_x^\infty dy \int_y^\infty dz\, P(z) = \int_x^\infty dz \int_x^z dy\, P(z) = \int_x^\infty dz\, (z-x) P(z). \quad (5)$$

Indeed, $\overline{h}(x) = \langle x \rangle\, \overline{F}(x) - x\overline{C}(x)$ and $\overline{h}(0) = \langle x \rangle$.

We have the derivatives: $P(x) = -d\overline{C}/dx$, $xP(x) = -\langle x \rangle\, d\overline{F}/dx$ and therefore $x/\langle x \rangle = d\overline{F}/d\overline{C}$. Also $\overline{C} = -d\overline{h}/dx$.

☺ ☻

$$G = \int_0^\infty dx\, \overline{F}(x)\, P(x) - \int_0^\infty dx\, \frac{x}{\langle x \rangle} \overline{C}(x)\, P(x) = \int_0^1 \overline{F} d\overline{C} - \int_0^1 \overline{C} d\overline{F}. \quad (6)$$

GINI

expressed via cumulative population

Using that $P(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}\overline{h}$ and $\overline{C}(x) = -d\overline{h}/dx$, we integrate by parts

$$\langle x \rangle \; G \; = \int_0^\infty dx \, \overline{h} \, \frac{\mathrm{d}^2 \overline{h}}{\mathrm{d}x^2} \; = \overline{h}(0)\overline{C}(0) - \int_0^\infty dx \, \overline{C}^2(x). \tag{7}$$

Replacing the boundary conditions we arrive at

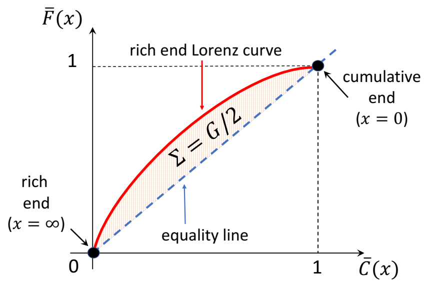
☺ ☻

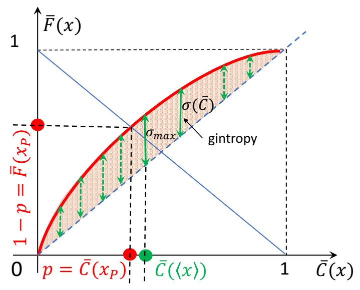
$$G \; = \; 1 \; - \; \frac{1}{\langle x \rangle} \int_0^\infty dx \, \overline{C}^2(x) \; = \; \frac{1}{\langle x \rangle} \int_0^\infty dx \, \overline{C} \, (1 - \overline{C}). \tag{8}$$

For scaling $P(x) = \frac{1}{\langle x \rangle} f\left(\frac{x}{\langle x \rangle}\right)$, $\overline{C}$ and G do not depend directly on $\langle x \rangle$.

Lorenz curve and gintropy

$$G = \int_0^1 (\overline{F} d\overline{C} - \overline{C} d\overline{F}) \qquad \int_0^1 \overline{F} d\overline{C} = 1/2 + \Sigma, \qquad \int_0^1 \overline{C} d\overline{F} = 1/2 - \Sigma.$$





GINI

expressed via gintropy: $G = 2\int_0^1 \sigma(\overline{C})\, d\overline{C}, \quad \Sigma = \int_0^1 \overline{F} d\overline{C} - 1/2 = \int_0^1 (\overline{F} - \overline{C}) d\overline{C}.$

It can be shown that the half-moon area,

$$\Sigma \equiv \int_0^1 \sigma(\overline{C})\, d\overline{C} \equiv \int_0^1 \left(\overline{F}(x) - \overline{C}(x)\right) d\overline{C}, \tag{9}$$

is exactly $G/2$. The integrand is alike an entropy-density we call it **gintropy**.

From the Lorentz curve geometry:

☺ ☻

$$\int_0^1 \sigma d\overline{C} = \int_0^1 \sigma d\overline{F} = \frac{1}{2} G. \tag{10}$$

Properties of gintropy

general

1. The gintropy is never negative: $\sigma = \overline{F} - \overline{C} = C - F \geq 0$ inspecting the integral

$$\sigma = \int_x^\infty \left(\frac{y}{\langle x \rangle} - 1 \right) P(y)dy = \int_0^x \left(1 - \frac{y}{\langle x \rangle} \right) P(y)dy \geq 0$$

   take first form for $y \geq x \geq \langle x \rangle$, the second form for the opposite case.

2. $\sigma(x)$ is maximal at $x = \langle x \rangle$: $d\sigma/dx = \left(1 - x/\langle x \rangle\right) P(x)$ changes sign there.

3. Convexity: $\frac{\mathrm{d}\sigma}{\mathrm{d}\overline{C}} = \frac{x}{\langle x \rangle} - 1$;

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}\overline{C}^2} = \frac{1}{\langle x \rangle} \frac{\mathrm{d}x}{\mathrm{d}\overline{C}} = -\frac{1}{\langle x \rangle P(x)} < 0.$$

4. For some PDF it looks like entropy density (see examples later).

GINI examples

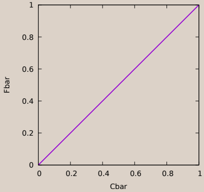
Communism: all incomes are equal

$P(x) \;=\; \delta(x-a)$ delivers $\langle x \rangle = a$, $\overline{C}(x) = \Theta(a-x)$ and $\overline{h}(x) = (a-x)\Theta(a-x)$.

This leads to $\overline{F} = (\overline{h} + x\overline{C}) / \langle x \rangle = \Theta(a-x)$ and by that

$$\sigma(x) \;=\; \overline{F}(x) \;-\; \overline{C}(x) \;=\; 0. \qquad (11)$$

As a conseqence also $G = 0$.



GINI examples

Communism++ : some of them are more equal. 50/50 – 100/0

Two-peak-PDF, $P(x) = p\delta(x-a) + (1-p)\delta(x-b)$, delivers $\langle x \rangle = pa + (1-p)b$.

$$\overline{C}(x) \;=\; p\Theta(a-x) + (1-p)\Theta(b-x) \qquad (12)$$

Having the value 1 for $x \le a$ and $(1-p)$ for $x \in [a,b]$, otherwise 0.

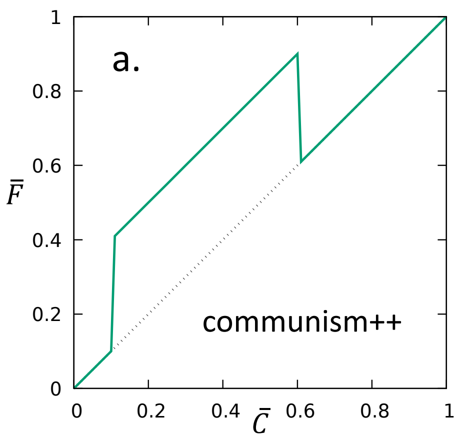
$$G \;=\; \frac{1}{\langle x \rangle} \int_a^b p(1-p)dx \;=\; \frac{(b-a)p(1-p)}{\langle x \rangle} \;=\; \frac{(\langle x \rangle - a)(b - \langle x \rangle)}{(b-a)\,\langle x \rangle}. \qquad (13)$$

Gintropy is a box: $\sigma(\overline{C}) \;=\; G\,\Theta(\overline{C}(b) \le \overline{C} \le \overline{C}(a))$.

$\overline{C}(b) = \frac{1-p}{2}$, $\overline{C}(a) = 1 - \frac{p}{2}$. **Pareto fraction:** $\overline{C}_P = \frac{1-G}{2}$, $\overline{F}_P = \frac{1+G}{2}$, $G_{\max}(p) = \frac{\sqrt{b}-\sqrt{a}}{\sqrt{b}+\sqrt{a}}$.

Lorenz curve examples

Communism++ : some of them are more equal.



GINI examples

Eco-window: 50/50 – 62/38

The PDF has the form: $P(x) = \frac{1}{b-a}\Theta(a \leq x \leq b)$. Then the cumulative

$$\overline{C}(x) = \begin{cases} 1 & (x < a) \\ \frac{b-x}{b-a} & \text{otherwise} \\ 0 & (b < x) \end{cases} \tag{14}$$

Obviously $\langle x \rangle = (a+b)/2$ and the GINI

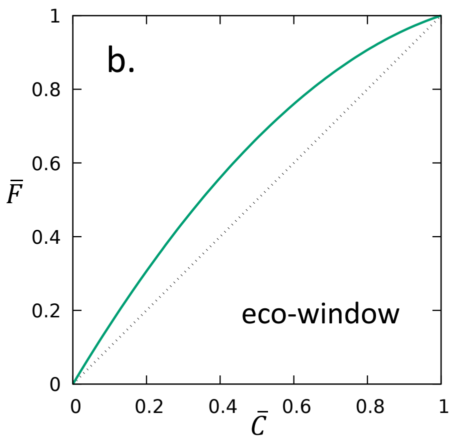
$$G = 1 - \frac{1}{\langle x \rangle}\left[a + \int_a^b \frac{(b-x)^2}{(b-a)^2} dx \right] = \frac{1}{3}\frac{b-a}{b+a}. \tag{15}$$

gintropy: $\sigma(\overline{C}) = \frac{b-a}{b+a}\overline{C}(1-\overline{C})$.

For $a = 0$: $G = 1/3$, $\sigma = \overline{C}(1-\overline{C})$ and the Pareto ratio: $\overline{F}_P/\overline{C}_P = 62/38$

Lorenz curve examples

Eco-window : equal chance between min and max.



GINI examples

Natural

68/32

The PDF scales: $P(x) \;=\; \frac{1}{\langle x \rangle} \mathrm{e}^{-x/\langle x \rangle}$.

The corresponding tail-cumulative probability: $\overline{C}(x) \;=\; \mathrm{e}^{-x/\langle x \rangle}$. The GINI becomes

$$G \;=\; 1 \;-\; \frac{1}{\langle x \rangle} \int_0^\infty \mathrm{e}^{-2x/\langle x \rangle} \, dx \;=\; \frac{1}{2}. \tag{16}$$

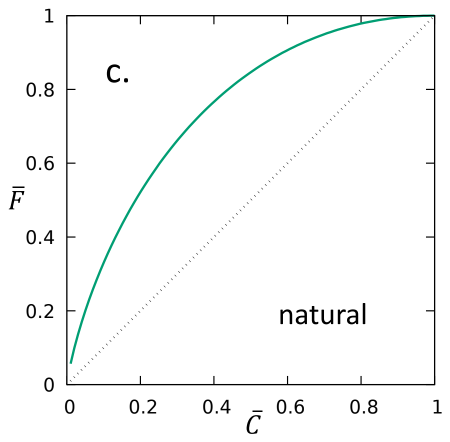
Entropy formula is constructed as follows:

$$\overline{h} \;=\; \int_x^\infty \mathrm{e}^{-y/\langle x \rangle} \, dy \;=\; \langle x \rangle \; \mathrm{e}^{-x/\langle x \rangle}. \tag{17}$$

$\langle x \rangle \, \overline{F} = \overline{h} + x\overline{C} = (x + \langle x \rangle)\mathrm{e}^{-x/\langle x \rangle}$ and $\sigma = \frac{x}{\langle x \rangle} \mathrm{e}^{-x/\langle x \rangle}$ $\quad\sigma \;=\; -\overline{C} \ln \overline{C}$ **!**

Lorenz curve examples

Natural : exponential PDF.



GINI examples

Capitalism: the rich gets richer

Start: $\overline{C}(x) = (1 + Ax)^{-B-1}$ leads to $\overline{h}(x) = \frac{1}{AB}(1 + Ax)^{-B}$. From this $\langle x \rangle = \overline{h}(0) = 1/AB$.

GINI becomes

$$G = 1 - AB \int_0^\infty (1 + Ax)^{-2B-2} dx = \frac{B+1}{2B+1}. \qquad (18)$$

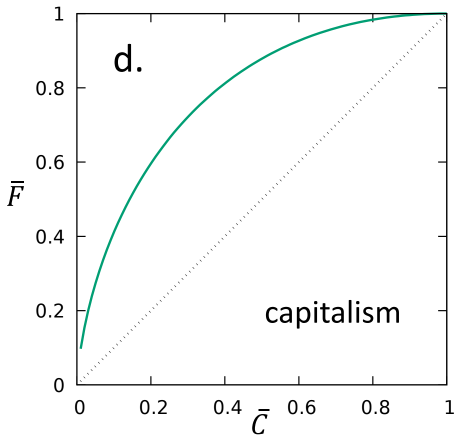
The gintropy

$$\sigma = Ax(B+1)(1+Ax)^{-B-1} = (B+1)\left(\overline{C}^{\frac{B}{B+1}} - \overline{C}\right). \qquad (19)$$

With $q = \frac{B}{B+1}$ we have $\sigma(\overline{C}) = \frac{1}{1-q}(\overline{C}^q - \overline{C})$ and $G = \frac{1}{q+1}$.

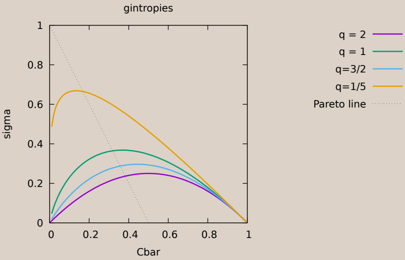
Lorenz curve examples

Capitalism : linear preference for the rich.



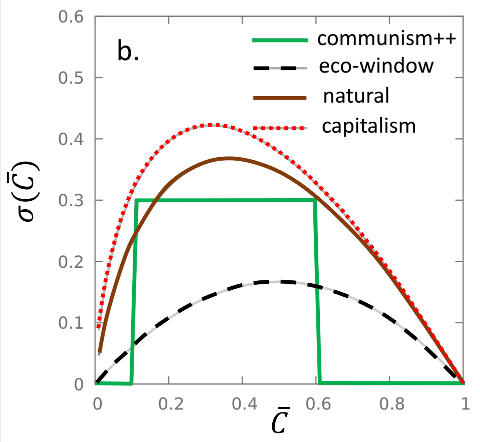
Gintropy

demonstrations



Gintropy

Comparison of examples



"An $e^{-\varepsilon/T}$ fit to ε-data justifies a thermal model"

is a FALLACY

Non-equilibrium stationary PDF, superstatistics

Non-thermal exponential PDF

with $T = E/\langle n \rangle$.

$$\rho(\varepsilon) = \left\langle \frac{\Omega_n(E-\varepsilon)}{\Omega_n(E)} \right\rangle = \left\langle \left(1 - \frac{\varepsilon}{E}\right)^n \right\rangle. \quad (20)$$

Statistics over several events:

$$\rho(\varepsilon) = \sum_{n=0}^{\infty} \left(1 - \frac{\varepsilon}{E}\right)^n P_n(E). \quad (21)$$

delivers the characteristic function of the PDF.

For $P_n(E) = a^n e^{-a}/n!$, Poissonian, $\langle n \rangle = a$ and

$$\rho(\varepsilon) = e^{-\frac{\langle n \rangle}{E}\varepsilon} \quad (22)$$

Non-thermal Tsallis–Pareto PDF

with $T = E/\langle n \rangle$ and $q = 1 + 1/(k+1)$.

Statistics over several events:

$$\rho(\varepsilon) = \sum_{n=0}^{\infty} \left(1 - \frac{\varepsilon}{E}\right)^n P_n(E). \qquad (23)$$

delivers the characteristic function of the PDF.

For NBD, $P_n(E) = \binom{k}{n} f^n (1+f)^{-n-k-1}$, one obtains

$$\rho(\varepsilon) = \left(1 + \frac{\langle n \rangle}{k+1} \frac{\varepsilon}{E}\right). \qquad (24)$$

Phase Space and Entropy

Boltzmann – Gibbs – Planck – Einstein

Volume in n dimensions and linear size E: $\Omega_n(E) = e^{S_n(E)}$.

$$\rho(\varepsilon) = \left\langle e^{S_n(E-\varepsilon)-S_n(E)} \right\rangle = \left\langle e^{-\varepsilon S_n'(E)+\frac{\varepsilon^2}{2}S_n''(E)+\ldots} \right\rangle$$

$$= \left\langle 1 - \varepsilon S_n' + \frac{\varepsilon^2}{2}\left((S_n')^2 + S_n''\right) + \ldots \right\rangle \qquad (25)$$

Interpretation of Tsallis parameters

Compare it with

$$\left(1 + (q-1)\frac{\varepsilon}{T}\right)^{-\frac{1}{q-1}} = 1 - \frac{\varepsilon}{T} + q\,\frac{\varepsilon^2}{2T^2} + \ldots \tag{26}$$

interprets

$$\frac{1}{T} = \langle S_n' \rangle = \langle \beta_n \rangle, \tag{27}$$

and

$$q = T^2 \left\langle (S_n')^2 + S_n'' \right\rangle = T^2 \left\langle \beta_n^2 \right\rangle - \frac{\mathrm{d}T}{\mathrm{d}E}. \tag{28}$$

Finally $$q = 1 + \frac{\Delta\beta_n^2}{\langle\beta_n\rangle^2} - \frac{1}{C}.$$

Construct another entropy

so that $S \to K(S)$ $\quad$ $K(S_{Renyi}) = S_{Tsallis}$

$$\rho_K(\varepsilon) = \left\langle e^{K(S(E-\varepsilon))-K(S(E))} \right\rangle = 1 - \varepsilon K_1 + \frac{\varepsilon^2}{2} K_2 + \ldots \tag{29}$$

Comparison with Tsallis–Pareto PDF delivers

$$\frac{1}{T_K} = \langle\beta\rangle\, K'; \qquad \frac{q_K}{T_k^2} = (K'' + (K')^2) \left\langle \beta^2 \right\rangle - K' \frac{\langle\beta\rangle^2}{C}. \tag{30}$$

e.g. $K(S) = \frac{e^{(1-q)S}-1}{1-q}$ leads to Tsallis and Renyi entropies ($\Delta\beta^2 = 0$).

Find $K(S)$ for $q_K = 1$, and then $K(S) = \sum_i p_i K(-\ln p_i)$.

LGGR model

local growth global reset

Transitions: from n to $n+1$ (local) and from any n to 0 (global).

$$\frac{\partial}{\partial t}P = -\frac{\partial}{\partial x}(\mu P) - \gamma P. \tag{31}$$

Stationary PDF-s:

- γ and μ rates constant:

$$P(\infty, x) = Q(0) \exp -\frac{\gamma}{\mu}x$$

- γ constant $\mu(x) = \sigma(x+b)$ linear:

$$P(\infty, x) = Q(0) \left(1 + \frac{x}{b}\right)^{-1-\gamma/\sigma}$$

Summary

- There are exp-s which are **non-thermal**.
- T stems from $E / \langle n \rangle$, q from $\Delta n^2 / \langle x \rangle$.
- Entropy mapping $S \to K(S)$ to ensure $q_K = 1$.
- Several $K(S) = \sum_i p_i K(-\ln p_i)$ formulas are possible.
- Gini index $\to$ gintropy reconstructs entropy formulas

REFERENCES

References

LGGR model

1. **Income data analysis**
   Z. Néda, I. Gere, T. S. Biró, G. Toth, N. Derzsy: *Scaling in Income Inequalities and its Dynamical Origin* Physica A (2020) 124491
2. **Citation analysis**
   Z. Néda, L. Varga, T. S. Biró: *Science and Facebook: the same popularity law!* PLOS ONE **12** (2017) 0179656
3. **Review of LGGR model**
   T. S. Biró, Z. Néda: *Unidirectional random growth with resetting* Physica A **499** (2018) 335
4. **Random, beyond linear master eq.**
   T. S. Biró, Z. Néda, A. Telcs: *Entropic Divergence and Entropy Related to Nonlinear Master Equations* Entropy **21** (2019) 993
5. **Thermal Power-Law**
   T. S. Biró, P. Ván, G. G. Barnaföldi, K. Ürmössy: *Statistical power law due to reservoir fluctuations and the universal thermostat independence principle* Entropy **16** (2014) 6497-6514

Bibliographical Notes

Networks and citation dynamics

1. **Tsallis/Pareto degree distribution in huge networks**
   Stefan Thurner, Fragiskos Kyriakopoulos, Constantino Tsallis: *Unified model for network dynamics exhibiting nonextensive statistics* Phys. Rev. E **76** (2007) 036111
2. **Avalanche Master Equation for Citations**
   A. Schubert, W. Glänzel: *A dynamic look at a class of skew distributions. A model with scientometric applications* Scientometrics **6** (1984) 149-167
3. **Waring distribution**
   J. O. Irwin: *The Place of Mathematics in Medical and Biological Statistics* J. Roy. Stat. Soc. A **126** (1963) 1-45
4. **Generalized Waring distribution**
   J. O. Irwin: *The Generalized Waring Distribution Applied to Accident Theory* J. Roy. Stat. Soc. A **131** (1968) 205-225
5. **Rate reconstruction from a truncated integral**
   W. Glänzel, A. Telcs, A. Schubert: *Characterization by Truncated Moments and Its Application to Pearson-Type Distributions* Z. Wahrscheinlichkeitstheorie verw. Gebiete **66** (1984) 173-183
6. **Avalanche Master eq. for networks**
   P. L. Krapivsky, S. Redner, F. Leyvraz: *Connectivity of Growing Random Networks* Phys. Rev. Lett. **85** (2000) 4629-4632
7. **Waring Degree Distribution in Network – eq.8 in**
   P. L. Krapivsky, G. J. Rodgers, S. Redner: *Degree Distributions of Growing Networks* Phys. Rev. Lett. **86** (2001) 5401-5404
8. **Hypergeometrical result in aging site network**
   S. N. Dorogovtsev, J. F. F. Mendes, *Evolution of networks with aging of sites* Phys. Rev. E **62** (2000) 1842-1845

Bibliographical notes

On citation statistics

9. **Citation universality**
   A. Chatterjee, A. Asim Ghosh, B. K. Chakrabarti, *Universality of citation distributions for academic institutions and journals* PLOS ONE **11** (2016) e0146762
10. **Article citations**
    E. S. Vierra, J. A. N. F. Gomes, *Citation to scientific articles: Its distribution and dependence on the article features* J. Infometrics **4** (2010) 1
11. **Citation of individuals**
    O. S. Garanina, M. Yu. Romanovsky, *Citation distribution of individual scientists: approximations of stretched exponential distribution with power law tails* arxiv: 1605.03741 (2016)
12. **Cut power law**
    C. Tsallis, M. P. de Albuquerque, *Are citations of scientific papers a case of nonextensivity?* EPJ B **13** (2000) 777
13. **Did the citation distribution change?**
    M. L. Wallace, V. Lariviére, Y. Gingras, *Modelling a Century of Citation Distribution* J. Infometrics **3** (2009) 296
14. **A century of citations**
    S. Redner, *Citation Statistics from 110 Years of Physical Review* Physics Today **58** (2005) 49
15. **The Zipf limit**
    Z. K. Silagadze, *Citations and the Zipf–Mandelbrot's law* arxiv: physics/9901035