

# New model of understanding income distribution through graduation of normalized Gini Mean Difference

Dmitry S. Schmerling  
[orcid.org/0000-0001-7659-2119](https://orcid.org/0000-0001-7659-2119)  
Professor  
Financial University, Moscow, Russia.  
Professor-Researcher  
National Research University Higher School of Economics,  
Moscow, Russia.

Devoted to Vladimir Igorevich Arnold  
(1937-2010)

## ABSTRACT

This work covers such a topical problem as the study of income and wealth inequality. It introduces the notion of “graduation”, i.e. rate of wage rate scale customary at time wage system. Graduation means introduction of the index  $m$  as a degree of the polynomial based on which the wage rate scale is distributed. In other words, using  $m$  allows to graduate Gini mean difference ( $0 < G < 1$ ) by assigning integral number or fraction  $m$ ,  $0 < m < \infty$  to every value of Gini index. Similarly, for each  $m$  we can calculate and estimate corresponding normalized Gini mean difference (the analog of Gini index used for simplicity of calculations).

Note that in this case we are not talking about distribution in real society, but in the simplest model of this society – a metaphoric community with only one person at each income level. This way we get the most distinct impression of the mechanism of income distribution that gives us the chance to assess scale of inequality. To some extent it clarifies the wide-spread Gini coefficient and “the nature of populations’ wealth”.

**JEL CLASSIFICATION CODES:** D31, C51

**KEY WORDS:** inequality, income distribution, Gini index, wage rate scale, new income distribution model.

## INTRODUCTION

This work covers such topical problem as income and wealth inequality in the context of professional competence. It introduces transformation of Gini index, reflecting income inequality, into index  $m$ , reflecting the wage rate scale customary in a certain country.

Traditional graph used for income inequality is the distance between Lorenz curve and the diagonal line meaning total equality. Instinctively we interpret this graph as “the closer to the diagonal, the less inequality, consequently, the better”. However, total equality usually is not

seen as ideal state of things. So, having only Gini index and Lorenz curve does not give us any understanding where the optimal curve is – the curve that would be considered fair for any reason.

In our model, we suggest that criterium for fair income inequality is consistency of income distribution with competence distribution. That is why we first need to transform classical Lorenz curve into the wage rate scale curve.

The model implies splitting continuous scale of Gini index into several intervals, ‘graduation’, and assigning to each interval a discrete value of index  $m$  expressed by integral number or fraction. Index  $m$  is the degree of polynomial on the basis of which the wage rate scale is distributed.

Note that in this case we are not talking about distribution in real society but about distribution in the simplest model of this society – a metaphoric community with only one person at each income level. This way we get the most distinct impression of the mechanism of income distribution among competence levels that gives us the chance to assess whether this income inequality is ‘objective’ and ‘fair’ inside our model, i.e. consistent with competence inequality, or not.

### GRADUATION OF GINI INDEX

For the sake of easier logic of calculations in this paper we split into intervals not classical Gini index that supposes Gini mean difference divided by  $\mu$  (average salary) but Gini *mean difference* divided by *maximum difference*. It turns out that these two approaches to normalization in our case give the same values, so the results of our approach are applicable to the officially published values of Gini index.

Income distribution problem has been known at least since M. Lorenz’s paper about concentration of wealth (Lorenz M.O., 1905)<sup>1</sup>, e.g. refer to Kendall, Stuart, Ord, 1977-1983, §§ 2.25<sup>2</sup>. There is wide range of literature on the problem of income, wealth and personal property differentiation of population, enterprises and territory. Gini index is widely used in the problem of income distribution inequality (Corrado Gini, 1912).

Let

$$\Delta_1 = \frac{1}{N(N-1)} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |x_j - x_k| f(x_j) f(x_k), j \neq k \quad (1)$$

(for discrete distribution), where  $x_1, x_2, \dots$  - value of income,  $f(x_1), f(x_2)$  – frequency of people with income  $x_1, x_2, \dots$  respectively.

Here  $\Delta_1^*$  is the normalized value of  $\Delta_1$ . So,  $\Delta_1^* = \Delta_1 / \Delta_1^{\max} \in [0,1]$ . Bigger value of  $\Delta_1^*$  means bigger social inequality.

Besides, the area over Lorenz curve and under diagonal is equal to  $\Delta_1/4 \mu$ , where  $\mu = \int_{-\infty}^{\infty} (x - a) f(x) dx$ ,  $f(x)$  is distribution density, assigned  $a=0$ , (see §§ 2,3 Kendall, Stuart, Ord, 1977-1983).

Dissipation (scattering) curve is equal to

<sup>1</sup> Biography of M. Lorenz, see Marshall, Olkin (1980), refer to Lorenz\_curve.Wikipedia.html.

<sup>2</sup> Besides, inequality was mentioned in V.I. Lenin’s paper «Development of capitalism in Russia» (1899).

$$\phi(x) = \frac{1}{\mu_1} \int_{-\infty}^x xf(x)dx$$

i.e. «incomplete first distribution's moment» (for more info the mean difference computation see §§ 2,3 Kendall, Stuart, Ord, 1977-1983).

Gini coefficient has been calculated and published for most countries for decades. The methods of evaluation for grouped data can be found in Gastwirth (1972), Modorres, Gastwirth (2006)<sup>3</sup>. Some values of normalized Gini index can be found in the table 1<sup>4</sup> (due to CIA data), e.g., Norway 0.25 (2008), France 0.32 (2008), Russia 0.423 (2008), Nigeria 0.437 (2003), the USA 0.45 (2007), Mexico 0.482 (2008), Haiti 0.538 (2001), Sierra Leone 0.629 (1989), South Africa 0.65 (2005), Namibia 0.707 (2003)<sup>5</sup>. In Namibia normalized Gini index has reached up to 0.75 in past two decades.

Most people are probably used to this kind of data and can compare income distribution in different countries by simply ranking them: whether distribution in one country is more or less differentiated than in another country. However, Gini coefficient is difficult to interpret directly – without comparison with any other country. That is why it might be useful to offer another simple measure one-to-one correspondent to Gini index, which will help us better understand the “physical sense” of inequality and income distribution.

The wide range of literature exists about the negative effect of high Gini coefficient > 0.3 (refer to Handbook of income distribution (Atkinson, Bourguignon, 2000)). The value 0.3 (or 1/3) can be considered as average European Gini coefficient, whereas the average for Scandinavia is about 0.25 (or 1/4).

For simplicity of the model, we used the simple formula for normalization of Gini coefficient  $G'$  ( $0 \leq G' \leq 1$ ):  $G' = G/\max_x G$ , where  $\max_x G$  is taken by all possible levels of discrete income distributions.

Let's consider the Model P (“graduation of normalized Gini mean difference”):  $x_i$  is income of people referring to  $i$  level of hierarchy of an enterprise, population, territory, etc.,  $i=1, \dots, n$ . For population with the lowest income  $i=1$ , for population with the highest income  $i=n$ . Income at  $i$  level is equal to  $x_i = \text{Const } i^m$ ,  $m = 1, 2, 3, \dots, M > 0$ .

Theorem: Normalized Gini mean difference  $G'_m(n)$  for the P-model at asymptotically  $n \rightarrow \infty$  is equal to

$$G'_m(n) \rightarrow \frac{m}{m+2} \quad (1a)$$

Proof:

$$G'(n) = \frac{\sigma(n)}{\max_x \sigma(n)}, \quad (2)$$

<sup>3</sup> Morgan J. The Anatomy of Income Distribution // Rev.Econ.Statist. (MITpress), 1962, v.44, N 3, p. 270-283.

<sup>4</sup> List of Countries by income equality [http://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_income\\_equality](http://en.wikipedia.org/wiki/List_of_countries_by_income_equality)

<sup>5</sup> In Bolivia Gini index is 0.592, but decile rate is too high – 168.1 (!)

where maximum for every possible  $\{X_{(1)}, X_{(2)}, \dots, X_{(n)}\}$ , so

$$\sum_{1 \leq i \leq n} x_{(i)} = C(n), \quad (3)$$

$$" \sigma " (n) = \frac{2\sqrt{\pi}}{n(n-1)} \sum_{1 \leq i \leq n} \left(i - \frac{n+1}{2}\right) x_{(i)}, \quad (4)$$

$X_{(i)}$  – i-order statistics<sup>6</sup>

$$G(n) = \frac{1}{n(n-1)} \sum_{1 \leq i, j \leq n} |x_i - x_j|. \quad (6)$$

Note that "σ"(n) and "σ" are the same notions, just when we say "σ"(n) we want to emphasize that "σ" depends on n; while when we say "σ" we don't consider this dependence on n important enough to mention.

Also note that we use "σ" in quotes to denote the special dispersion of Gini mean square difference. It was proposed by David (1968, 1981), apparently, to emphasize *the similarity* (not identity) with standard deviation.

Now let's use another expression for "σ":

$$" \sigma " = \frac{\tilde{\kappa} 2 * \sqrt{\pi}}{n(n-1)} \left\{ \sum_1^n i x_{(i)} - \frac{n+1}{2} \sum_1^n x_{(i)} \right\} \quad (7)$$

and calculating

$$\max_x " \sigma " = \frac{\sqrt{\pi} * \tilde{\kappa}}{n} S_m(n) \quad (8)$$

<sup>6</sup> See David (1981), (7.4.1.) and §7.4., ex. 7.4.1., §9.6, (9.6.1.), where asymptotical "σ" is discussed. So

$E " \sigma " = 2 \sqrt{\pi \int_{-\infty}^{+\infty} x \left[ P_{(x)} - \frac{1}{2} \right] dP_{(x)}}$ , "σ" – unbiased estimator for normal samples,  $P_{(x)}$  – cumulative distribution

function.

where  $S_m(n) = \sum_{k=0}^{n-1} k^m$  <sup>7</sup>

i.e., the sum of integers raising to power  $m=1, 2, 3, \dots$ ,  $k = 0, 1, 2, \dots, n-1$ , allows us to have a formula for (3).

So, let

$$" \sigma " = \frac{2\tilde{\kappa}\sqrt{\pi}}{n(n-1)} \left\{ \tilde{S}_{m+1}(n) - \frac{n+1}{2} \tilde{S}_m(n) \right\} / \frac{\tilde{\kappa}\sqrt{\pi}}{n} \tilde{S}_m(n) \quad (9)$$

Here m means the degree of polynomial. Therefore:

$$G'_m(n) = \frac{2}{n-1} \left\{ \frac{\tilde{S}_{m+1}(n)}{\tilde{S}_m(n)} - \frac{n+1}{2} \right\} \quad (10)$$

According to Graham, Knuth, Patashnik (1994), (61.78),

$$\tilde{S}_m(n) = \frac{1}{m+1} \sum_{0 \leq k \leq n} \binom{m+1}{k} B_k n^{m+1-k} \quad (11)$$

where  $B_k$ ,  $k = 0, 1, 2, \dots$ , (Jacob) Bernoulli's number, that is

k	1	2	3	4	5	6	7	8	9	10	11	12	...
$B_k$	1	$-\frac{1}{2}$	0	$-\frac{1}{30}$	0	$\frac{1}{42}$	0	$-\frac{1}{30}$	0	$\frac{5}{66}$	0	$-\frac{691}{2730}$	...

Formula  $S_m(n)$ ,  $m = 0, 1, 2, \dots, 10$  see the above mentioned book p.314.

Formula (10) at  $n \rightarrow \infty$  implies that:

$$G'_m(n) \cong \frac{2}{n-1} \left\{ \frac{m+1}{m+2} n - \frac{n+1}{2} \right\} \approx \frac{m}{m+2}, QED.$$

for  $G'_m(n)$ <sup>8</sup>

m	1	2	3	4	5	6	7	8	9	10	...
$G'_m(n)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{3}{5}$	$\frac{2}{3}$	$\frac{5}{7}$	$\frac{3}{4}$	$\frac{7}{9}$	$\frac{4}{5}$	$\frac{9}{11}$	$\frac{5}{6}$	...

Let's revert to the beginning of the summary. If P is equal to Const=1, so  $G'=0,25$  (Norway) means  $m < 1$  (fraction),  $G'=0,327$  (France) means  $m \approx 1$ ,  $G'=0,423$  (Russia) means  $1 < m < 2$ ,  $G'=0,482$  (Mexico) provides  $m \approx 2$ ,  $G'=0,538$  (Haiti) -  $2 < m < 3$ , Sierra Leone with 0,629 leads to  $3 < m < 4$ , Republic of

<sup>7</sup> See Graham, Knuth, Patashnik (1994)

<sup>8</sup> At  $m=1$  - Gini index is accurate (definite), at every  $n = 1, 2, 3, \dots$

South Africa with 0,65 provides m more, but Namibia, according to different sources, 0,707 < G' < 0.75, can have pretensions to 5 < m < 6.

## CONCLUSION

Companies, communities, regions and countries can be relatively divided into linear (m=1), quadratic (m=2), cubic (m=3), «tetradic» (m=4), «pental-power» (m=5), «hexal» (m=6), etc. For example, in Moscow, where G' raised to 0.62 in different years it can be assumed cubic type of income distribution. But we should keep in mind that the true value of G' in Moscow can be equal to 0.60-0.70 according to many specialists.

As for Russian companies, the data of some large companies which are regularly discussed by 'Vedomosti' newspaper show the following results<sup>9</sup>. The bonuses of administration personnel (exclusive of salary) compiles 1-3 million USD per year compared to salary of senior staff at the level i=1 compiles 1.5-3.0 thousand USD and bonus which is equal to 10 – 30 thousand USD altogether form the value 2 < m < 3. m can probably reach 4, as in 2010 the General Manager's salary of the leading companies in Russia can reach 25 million USD per year<sup>10</sup>.

Kiruta A. Ya. analyzed the discussed here problem in terms of Lorenz curve and sent the 8-paged comment to the author on May, 3<sup>rd</sup>, 2011, at 2.34 pm.

It should be mentioned that it is also possible to consider real values (due to interpolation)  
 $m = 2G^{\wedge} / (1 - G^{\wedge})$

Here G^ is normalized value of normalized Gini mean difference. In Moscow in 2009 G^=0.521, that gives m=2.742. For Norway (2008) G^=0.25 and m=0.667.

This needs some clarification. The model discussed above can correlate to traditional static distribution. For Pareto distribution with the same G^, as in our P-model, m < 1, ensures finite variance (dispersion), but for log-logistic distribution m < 2 is required. As for log-normal distribution variance (dispersion) doesn't approach infinity, but it is increasing slowly on condition that m is slowly increasing as well. It should be mentioned that Pareto and log-logistic distributions describe right (top) tail but log-normal ones - little income and describe poorly right tail.

At high degree of inequality in the P-model little (rich) part of the society strives to increase the income so that the high distribution tails are getting heavy and variance (dispersion) tends to infinity. At the same time, the part of the society with average income reacts slowly at the increase of m.

To sum up, the discussed approach to income inequality research touches many aspects of social activities and dynamics and deserves further exploration.

---

<sup>9</sup> See master degree thesis: Milek O. The study of income distribution with the help of heavy tail distribution. – M., HSE, 2010 (in Russian)

<sup>10</sup> See [http://www.rfcor.ru/print/news\\_rfc\\_987htm](http://www.rfcor.ru/print/news_rfc_987htm)

## REFERENCES

1. Graham, Ronald L.; Knuth, Donald E.; Patashnik, Oren (1994). *Concrete mathematics: A foundation for computer science* (Second ed.). Reading, MA: Addison-Wesley Publishing Company. pp. xiv+657. ISBN 0-201-55802-5. MR 1397498.
2. David, H. A. (1970). *Order Statistics*. Wiley, New York. (2nd ed., 1981)
3. Mathematical Reviews (MathSciNet): MR597893
4. Zentralblatt MATH: 0553.62046
5. Sir Maurice Kendall, Alan Stuart and J. Keith Ord. *The advanced theory of statistics / 4th ed.* London : C. Griffin, 1977-1983. §§2.21-2.26
6. *Inequalities: Theory of Majorization and Its Applications* (1980) Albert W. Marshall, Ingram Olkin, Academic Press, ISBN 978-0-12-473750-1
7. Atkinson A.B., Bourguignon F., eds. *Handbook of Income Distribution*, vol. I .- S.I.: Elsevier, 2000.
8. Chakravarty S.R. *Ethical Social Index Numbers*. – N.Y.: Springer-Verlag, 1990.
9. David H.A. Gini's Mean Difference Rediscovered // *Biometrika*, 1968, v. 55, p. 573-575.
10. Dikhanov Y. *Decomposition of Inequality Based on Incomplete Information*. – Washington, D.C.: The World Bank, 1996. – 26 pp. (A contrib. paper to the LARIW 24<sup>th</sup> Gen.Conf. Lillehammer, Norway, aug.18-24, 1996.
11. Foster J.F., Sen Amartya. *On Economic Inequality*, exp. ed. With Substantial Annex. Oxford: Oxford Univ. Press, 1997. – 280 pp.
12. Gastwirth Y.L. The Estimation of the Lorenz Curve and Gini Index // *Rev. Econ. Statistics*, 1972, v. 52, #3, p. 306-316. doi: 10.2307/1937992.
13. [Gini\\_coefficient.wikipedia.htm](#)
14. Gini, Corrado. *Variabilità e mutabilità, contributo allo studio delle distribuzioni e relazioni statistiche* // *Studi Economico-Giuridici della R. Università di Cagliari*, Repr. In *Memorie di Metodologia Statistica / Piretti E, Salvemini T. ed.* – Rome: Libreria Eredi Virgilio Veschi, 1955. (The reprinted work of 1912).

15. Lorenz M.D. Methods of Measuring the Concentration of Wealth // Publ. Amer. Statist. Ass., 1905, v. 9, #70, p. 209-219.
16. Moderres R., Gastwirth J.L. A Cautionary Note on Estimating the Standard Error of the Gini Index of Inequality // Oxford Bull. Econ. Statist., 2006, v. 68, p. 385-390. doi: 10.1111/j.1468-0084.2006.00167
17. [http://en.wikipedia.org/wiki/list\\_of\\_countries\\_by\\_income\\_equality](http://en.wikipedia.org/wiki/list_of_countries_by_income_equality)
18. Kleiber C., Kotz S. Statistical Size Distributions in Economics and Actuarial Sciences. – Hoboken, NJ: Wiley, 2003. – 332p.
19. Dagum C. Income inequality measures. – In: Encyclopedia of Statistical Sciences. 16 vols. N.L.Johnson, S.Kotz ed-in-chiefs. – N.Y. e.a.: Wiley, 2006. -9686 pp., p.3387-3405.