METHODS OF DECOMPOSING INEQUALITY MEASURES: A REVIEW ARTICLE

Noriyuki Takayama

武蔵大学論集第二十七巻第三・四・五号抜刷
昭　和　五　十　四　年　十　二　月

METHODS OF DECOMPOSING INEQUALITY MEASURES: A REVIEW ARTICLE

Noriyuki Takayama

What governs the inequality of size distribution of income and wealth? This is the most difficult question in the study on income distribution and yet has not been answered persuasively by economists. The decomposition of summary indices into relevant components is one way for us to find predominant factors of inequality. This paper makes a brief review of three methods of the decompsition.

1. *Decomposition by Sub-Populations*

1.1.1 (Kuznets' Suggestions) In an illuminating paper, S. Kuznets (1955) pointed out a so-called inverted-U shape pattern of income inequality change in the course of a country's economic growth. He suggested three sets of factors to explain this change; intersector differences in per capita income, intrasector distributions, and sector weights (see his paper pp. 12~18). With the help of a numerical illustration, he succeeded in bringing out the implications of the population shift from agricultural to non-agricultural sectors (see the fifth conclusion on page 15, in particular).

* This is an extended version of Takayama (1978) submitted to the first CAMS-Hitotsubashi Seminar on Income Distribution by Sectors and Overtime in East and Southeast Asian Countries held at Narita, Japan, in September 1977.

1.1.2 (Theil's Method) H. Theil (1967) has shown that his measure of inequality T based on entropy can be neatly decomposed into inequality within groups Tw_i and inequality between them T_b.

$$T=T_b+\sum w_i Tw_i,$$

where w_i is the weight and is equal to the income share of i-th group.

It must be remembered that T_b is defined as:

$$T_b=\sum[(\mu_i/\mu)\, log\, (\mu_i/\mu)]\, g\, (\mu_i),$$

where μ, μ_i, and $g(\mu_i)$ are the mean income of the whole population, the mean income of the i-th group, and its population share, respectively. T_b is calculated with group *means* μ_i only and is independent of *the distribution* within groups, that is, T_b is the income inequality when all the income distributions within groups are *assumed* to be perfectly equal ($Tw_i=0$ for all i).

It is obvious that this decomposition becomes more meaningful when the between-group inequality T_b is relatively large. If T_b as a percentage of T is considerably large, then intergroup differences in per capita income (the first factor suggested by Kuznets) can account for an overwhelming part of total inequality.

1.1.3 (Toyoda's Generalization) Takashi Toyoda (1975) extends Theil's method of decomposition to his measures of the expected-utility type B which are made up under a general weighting system with one parameter (α). When α equals unity, his measure is reduced to Theil's mearure T, and when $\alpha=2$, his measure equals half the square of coefficient of variation. Furthermore, when α is less than unity, his measure corresponds to Atkinson's measure A in following formula (see Atkinson (1970)):

$$A=1-(1-\alpha B)^{1/\alpha};\ \alpha\neq 0,\ \alpha<1,$$
$$=1-exp\,(-B);\ \alpha=0.$$

In this sense, B can be treated as Atkinson's variant when $\alpha<1$ (remark: $\alpha=1-\varepsilon$).

$$B=(1/\alpha)\,[1-\sum(y_k/\mu)^{\alpha}f(y_k)]\,;\ \alpha\neq0,\ \alpha<1,$$
$$=-\sum[\log(y_k/\mu)]f(y_k)\,;\ \alpha=0,$$
$$=\sum[(y_k/\mu)\log(y_k/\mu)]f(y_k)\,;\ \alpha=1,$$
$$=(1/\alpha)\,[\sum(y_k/\mu)^{\alpha}f(y_k)-1]\,;\ \alpha>1,$$

where y_k, $f(y_k)$ are the income of the k-th class and its population share.

Corresponding to the form of decomposition of Theil's measure, we can decompose Toyoda's measure as follows.

$$B=B_b+\sum w_iBw_i,\ \ w_i=g(\mu_i)(\mu_i/\mu)^{\alpha}.$$

When α equals unity or zero, we have $\sum w_i=1$ and the total within-group inequality $\sum w_i\,Bw_i$ becomes just the weighted average of the within-group inequality. The smaller the value of α, the heavier the weight given to the lower income classes at the margin. (Mizoguchi et al. (1977) fully utilizes the decomposition method of Toyoda's measure B in finding out main factors governing the inequality changes of income distribution in postwar Japan.)

1.1.4 (Decomposition of the Gini Coefficient in Theil's Sense) The Gini coefficient G can be decomposed into the between-group Gini G_b and the within-group Gini Gw_i if the decomposition is applied to the decile or quintile groups, that is, the strong assumption of income grouping in a non-decreasing order is necessary for the Gini coefficient to be decomposable in Theil's sense. The weight w_i of the Gini coefficient is given by:

(1) $\quad w_i=(\mu_i/\mu)\,(g(\mu_i))^2.$

This is equal to the income share of the i-th group multiplied by its population share.

This method of decomposition is applied to the measurement of poverty by Takayama (1979 b) in the censored income distribution truncated from above by the poverty line.

1.2 (Mangahas' Method) Mangahas (1975) presents a different method of decomposition of the Gini coefficient which is free from the strong assumption as is mentioned above.

$$G = L_b^* + \sum w_i^* G w_i.$$

The weight w_i^* attached to the within-group Gini is given by:

$$w_i^* = (\mu_i/\mu)\, g(\mu_i).$$

It is equal to the income share of the i-th group, which is different from (1).

L_b^* (the between-group component) is defined as:

$$L_b^* = \sum_{i>i} (D_{ij}/\mu)\, g(\mu_i)\, g(\mu_j),$$

where D_{ij} is the Gini difference.

$$D_{ij} = (h_i - h_j)'\, P\, (h_i - h_j).$$

Note that h_i (or h_j) is the $n \times 1$ vector of the population percentage (the density) of each income class in the i-th (or j-th) group, where n indicates the number of income classes. P is given by QY, where Q is an $n \times n$ matrix with ones on the diagonal, twos below it, and zeros elsewhere, and Y is an $n \times n$ diagonal matrix with y_k (the income received by families in income class k) as the typical diagonal element.

$$Q = \begin{vmatrix} 1 & 0 & 0 & . & . & . & . \\ 2 & 1 & 0 & . & . & . & . \\ 2 & 2 & 1 & 0 & . & . & . \\ . & . & 2 & 1 & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & 1 & 0 \\ . & . & . & . & . & 2 & 1 \end{vmatrix}, \quad Y = \begin{vmatrix} y_1 & 0 & 0 & . & . & . & . \\ 0 & y_2 & 0 & 0 & . & . & . \\ 0 & 0 & y_3 & 0 & 0 & . & . \\ 0 & 0 & 0 & y_4 & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & 0 \\ . & . & . & . & . & 0 & y_n \end{vmatrix}, \quad (h_i - h_j) = \begin{vmatrix} h_{i1} - h_{j1} \\ h_{i2} - h_{j2} \\ h_{i3} - h_{j3} \\ h_{i4} - h_{j4} \\ . \quad . \quad . \\ . \quad . \quad . \\ h_{in} - h_{jn} \end{vmatrix};$$

D_{ij} is *zero* if and only if both income distributions are *identical*, and therefore it is not zero if two unequal distributions have equal means. In this respect, D_{ij} compares twe group's size *distribution* of income, not merely their *means*. This is the main difference between L_b^* and G_b (the between-group Gini in Theil's sense), where G_b is given by:

$$G_b=\sum_{i>j}(|\mu_i-\mu_j|/\mu)\,g(\mu_i)\,g(\mu_j).$$

G_b is not equal to L_b^* in general, but they have the same value in a special case where $Gw_i=0$ for all i, actually. (With regard to other attempts at Gini disaggregation, see Bhattacharya-Mahalanobis (1967), Pyatt (1976), Rao (1969) and Soltow (1967).)

1.3 (Numerical Examples) The difference between L_b^* and G_b can be clarified by a numerical illustration (see Table I).

Table I

Case	Income Groups	Mangahas' Method					Theil's Method			Between/Total		
		Gw_i	$\sum w_i^* Gw_i$	L_b^*	$(./G)$	G	$\sum w_i Gw_i$	G_b	$(./G)$	T_b	A_b $(\varepsilon=1)$	A_b $(\varepsilon=2)$
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
I	(1, 5) (2, 4)	1/3 1/6	1/4	1/24	(1/7)	7/24	—	0	(0)	0	0	0
II	(2, 4) (6, 12)	1/6 1/6	1/6	1/6	(1/2)	1/3	1/12	1/4	(3/4)	0.698	0.721	0.75

In this illustration, we deal with a four-person model which has two groups (agricultural and nonagricultural sectors, for example). Each group is composed of two persons. In case I two groups have equal means (3, 3), but have different dispersions. On the contrary, in Case II two groups have equal-group inequality, but have different means (3, 9).

In Case I where intergroup difference in per capita income is zero, the between-group inequality in Theil's sense is zero (see cols. (7)~(11)), while the between-group component $L_b{}^*$ in the Mangahas version is 1/24 ($\neq 0$). This reflects that $L_b{}^*$ compares not only group means but also their intragroup inequalities. In addition, $L_b{}^*$ compares another element. In Case II intragroup inequalities are the same between two groups. Even in this Case $L_b{}^*$ ($\doteqdot 1/6$) is different from G_b ($=1/4$). In this respect, the between-group component of the Gini in the Mangahas version is really a residual term compounding many factors.

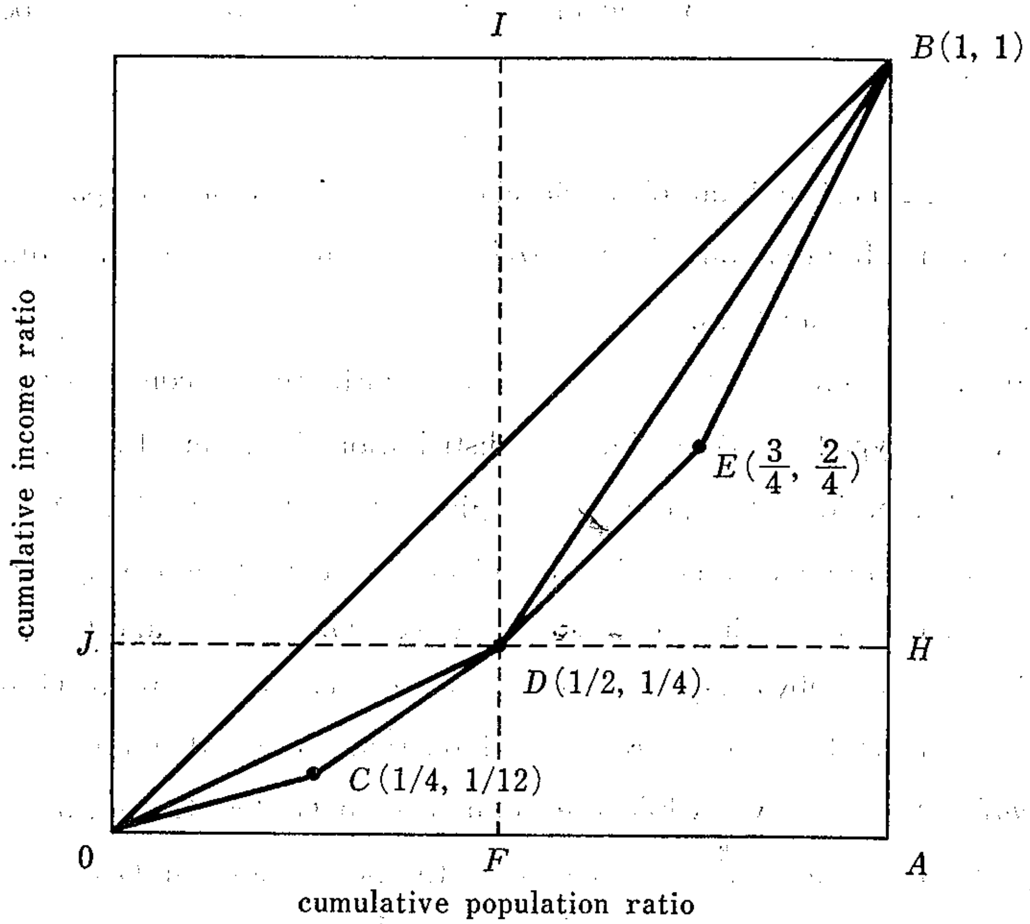
As Bhattacharya-Mahalanobis (1967, p. 150) suggests, it is reasonable to lay down that the between-group component should not change if the group distributions are changed, keeping all group means fixed. This is because we want to separate out the pure component of the intergroup difference in per capita income. For this purpose it is advisable not to adopt the Mangahas decomposition method but to use the Theil method.

Another point is worth noting. $L_b{}^*$ accounts for 1/2 of the total Gini, while G_b does 3/4 of it in Case II (see cols. (4) & (8) in Table I). Table I also presents figures of the between-group inequality as a percentage of total inequality using measures of expected-utility type in cols. (9)~(11) (A denotes Atkinson's measure). They tell us that the intergroup difference in per capita income may account for around 70%. This percentage corresponds to that of G_b. Then we can say $L_b{}^*$ (the between-group component of the Gini in the Mangahas version) *underestimates* the intergroup difference in per capita income.

In Case II two groups lie in different non-overlapping size-ranges, which satisfies the strong assumption for the Gini coefficient to be decomposable in Theil's sense. This enables us to present a *geometrical exposition* of the Gini disaggregation. See Diagram I. The Gini coefficient corresponds to area *OCDEB* divided by triangle *OAB*. The between-Gini is shown to equal triangle *ODB* divided by triangle *OAB*, while the weighted within-

Ginis ($w_i G w_i$) are equal to triangle OCD and DEB divided by triangle OAB. Note that the respective weights (w_i) are shown to correspond to area $OFDJ$ and $DHBI$.

DIAGRAM I



2. *Decomposition by Income Components*

2.1 (Rao's method) V. M. Rao (1969) demonstrates that the overall Gini G can be decomposed additively as:

(2) $G=\sum\theta_l \cdot \tilde{G}_l$,

where θ_l is the income share of income type l (wages and salaries, property income, wife's income, subsidies and gifts, etc.) and $\tilde{G}_l$ is the "pseudo" Gini coefficient computed from unordered income shares of type l. Note that $\tilde{G}_l$ is identical with G_l (the genuine Gini) if the rank correlation between the distribution of income type l and the total distribution is unity.

Formula (2) is quite useful when we want to know (a) what type of income dominates G, (b) what type of income has an equalizing effect on G, and (c) how changes in income shares of respective income types are related to changes in G.

2.2 (A Numerical Example) Special attention should be paid to the *pseudo* Gini. It often takes a *negative* value, which has a big equalizing effect on the overall Gini.

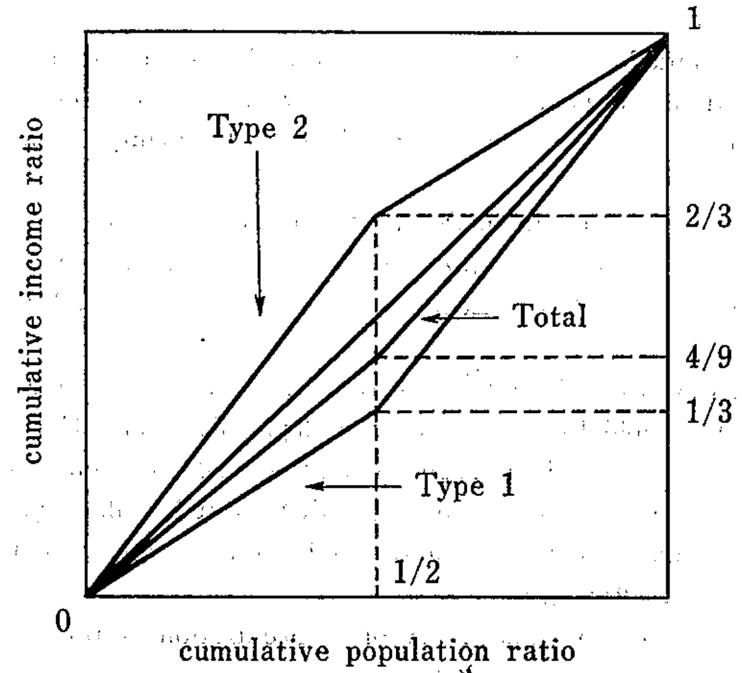
Table II presents a two-person model with two income components (type 1 and type 2). Total income distribution is given by (4, 5), and income distribution of type 1, 2 are given by (2, 4), (2, 1). Note that income distribution of type 2 is not given in a non-decreasing order of income, which results in a difference between $\tilde{G}_2$ (−1/6) and G_2 (1/6, the genuine Gini). Obviously the negative $\tilde{G}_2$ decreases the total Gini very much. It must be remembered, therefore, that the pseudo Gini $\tilde{G}_j$ *is worth computing* to be shown, while the genuine Gini G_j has less significance in computation as far as the comparison with the overall Gini G is concerned. (The importance of a negative $\tilde{G}_j$ is also shown in Diagram II, where the Lorenz curves of total income distribution and of income distribution type 1, 2 are given.)

The Gini disaggregation (2) is found to be useful also in the study on expenditure distribution.

Table II

Income Distribution	Total (4, 5)	Type 1 (2, 4)	Type 2 (2, 1)
The Ginis; $G, \tilde{G}_j$	1/18	1/6	−1/6
Income Shares	—	2/3	1/3

DIAGRAM II



3. *Decomposisition by Multiple Factors*

Wage income is defined as the wage rate multiplied by working hours. Wealth of land-holdings is given by the unit land-price multiplied by its area. These are examples of economic variables which have multiple factor components. It is very convenient for us to separate out the real element of economic difference from the nominal one.

The decomposition by multiple factors can be only given by Toyoda's measure B with $\alpha=0$.

(3) $B(y)=B(p)+B(x)$; $\alpha=0$, where $y=px$.

Note that the average price μ_p is defined as:

$$\mu_p=\sum y_k f(y_k)/[\sum x_k f(x_k)]=(\mu/\mu_x)$$

An application of formula (3) is found in Takayama (1979 a), for example.

References

ATKINSON, A.B. (1970), "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 244~263.

BHATTACHARYA, N. and MAHALANOBIS, B. (1967), "Regional Disparities in Household Consumption in India," *Journal of the American Statistical Association*, 62, 143~161.

KUZNETS, S. (1955), "Economic Growth and Income Inequality," *American Economic Review*, 45, 1~28.

MANGAHAS, M. (1975), "Income Inequality in the Philippines: A Decomposition Analysis," in JERC-CAMS, *Income Distribution, Employment, and Economic Development in Southeast and East Asia*, Japan Economic Research Center and Council for Asian Manpower Studies.

MIZOGUCHI, T., TAKAYAMA, N. and TERASAKI, Y. (1977), "Overtime Changes of Size Distribution of Household Income under the Rapid Economic Growth: the Japanese Experience," a paper presented to the 2nd Asian Regional Conference of the International Association for Research in Income and Wealth held at Manila, April 1977.

PYATT, G. (1976), "On the Interpretation and Disaggregation of Gini Coefficient," *Economic Journal*, 86, 243~255.

RAO, V.M. (1969), "Two Decompositions of Concentration Ratio," *Journal of Royal Statistical Society*, Secies A, Part 3, 132, 418~425.

SOLTOW, L. (1967), "The Distribution of Income Related to Changes in Distributions of Eduction, Age, and Occupation," *Review of Economics and Statistics*, 450~453.

TAKAYAMA, N. (1978), "A Note on the Decomposition of Inequality Measures," in OSHIMA, H.T. and MIZOGUCHI, T. eds., *Income Distribution by Sectors and Overtime in East and Southeast Asian Countries*, CAMS and Hitotsubashi University.

TAKAYAMA, N. (1979 a), "A Note on Wealth Distribution in Japan," Discussion Paper No. 2, Department of Economics, Musashi University (a paper presented to the 2nd CAMS-Hitotsubashi Seminar on Poverty Levels at Different Stages of Development in Asian Countries held at Sapporo, October 1978).

TAKAYAMA, N. (1979 b), "Poverty, Income Inequality, and Their Measures: Professor Sen's Axiomatic Approach Reconsidered," *Econometrica*, 47 (3).

THEIL, H. (1967), *Economics and Information Theory*, Rand McNally.

TOYODA, T. (1975), "Income Inequality: the Measures and Their Comparisons," *Kokumin Keizai*, 134, 15~41, in Japanese.