KYBERNETIKA — VOLUME *28* (1992), NUMBER 4, PAGES 325 – 332

# ON THE ASYMPTOTIC OPTIMUM ALLOCATION IN ESTIMATING INEQUALITY FROM COMPLETE DATA[1]

Mara Angeles Gil and Ignacio Martínez

Studies dealing with the quantification of inequality of a population with respect to a given quantitative attribute, provide us with a large class of measures. Among these, we can distinguish, because of their properties and operativeness, the ones coinciding with, or being ordinally equivalent to, the dimensionless "additively decomposable inequality indices".

As indicated in previous papers, many populations, whose inequality in relation with an attribute is useful to quantify, are too large to be censused but large samples from them can be drawn and they arise naturally stratified. On the basis of these last two advantages, we will approach in this paper the optimum allocation in estimating inequality, and a comparison with the proportional allocation, and with the absence of strata, will be later established.

## 1. INTRODUCTION

There are several quantitative attributes, many populations vary with respect to. The *inequality* of a population with respect to a given quantitative attribute (or variable) is understood as the population variation when the magnitude of attribute values are relevant for that variation (that is, they numerically indicate if those values are close or remote from each other). Three elements may then be taken into account in inequality measurement: the number of different values the attribute can take on in the population, the magnitude of those values, and the associated population distribution. The main practical purpose of inequality measurement is serving as a criterion (usually dimensionless) to compare populations (countries, enterprises, years, etc.).

Inequality and its numerical quantification is a topic having many interesting applications in fields like Economics (income inequality, wealth inequality, etc.; see, for instance, [9, 16]), Industry (industrial concentration; see, for instance [13]), and others.

Several inequality measures have been suggested in the literature, some of them being closely related to measures in Information Theory. Among these last ones, those coinciding with (or being an increasing function of) additively decomposable inequality indices are largely accepted. These indices have been quite recently introduced (cf. [4, 7, 8, 9, 15] and [17]) through different axiomatic approaches.

The behaviour of some of the additively decomposable inequality indices in stratified random sampling from complete data has been analyzed in previous studies ([4], [5], and [10]). Our purpose is now complementing last three studies by approaching the optimum allocation in estimating inequality in stratified sampling from complete data. This optimum allocation will be determined on the basis of one of the following objectives: either maximizing the precision of estimation or minimizing the size of the sample.

[1] The research in this paper was supported in part by DGICYT Grants No. PS89-0169 and No. PS89-0170. Their financial support is gratefully acknowledged.

## 2. ASYMPTOTIC BEHAVIOUR OF ADDITIVELY DECOMPOSABLE INEQUALITY INDICES IN STRATIFIED SAMPLING

Consider a finite uncensused population of $N$ individuals which is divided into $r$ non-overlapping strata. Assume that each individual value with respect to a given attribute is positive, $x_1^\star, \dots, x_M^\star$ being the possible different attribute values in the population ($x_i^\star > 0$). Let $N_k$ be the number of individuals in the $k$th stratum, $W_k = N_k/N$, and let $p_{ik}$ and $p_{i.}$ respectively denote the probabilities that a randomly selected individual in the $k$th stratum and in the whole population has an attribute value equal to $x_i^\star$ ($i = 1,$
$\dots, M,\ k = 1, \dots, r$).

Then, if $\mathbf{X}^\star$ denotes a random variable whose distribution in the $k$th stratum coincides with that of a random variable $X_k^\star$ taking on values $x_1^\star, \dots, x_M^\star$ with respective probabilities $p_{1k}, \dots, p_{Mk}$ ($k = 1, \dots, r$), then the inequality in the population with respect to the given attribute may be quantified by means of the measures below.

**Definition 2.1.** The measure $\mathbf{I}^\beta$ ($\beta \in \mathbb{R}$) associating with $\mathbf{X}^\star$ the value

$$\mathbf{I}^\beta(\mathbf{X}^\star) = \sum_{i=1}^{M} p_{i.}\, \phi_\beta\left(\frac{x_i^\star}{\mathsf{E}(\mathbf{X}^\star)}\right) = \sum_{i=1}^{M}\sum_{k=1}^{r} p_{ik}\, \phi_\beta\left(\frac{x_i^\star}{\mathsf{E}(\mathbf{X}^\star)}\right)$$

(where $\mathsf{E}(\mathbf{X}^\star) = \sum_i p_{i.}\, x_i^\star = \sum_i \sum_k p_{ik}\, x_i^\star$), $\phi_\beta$ being the real-valued function given by

$$\phi_\beta(x) = \begin{cases} x^\beta - 1 & \text{if } \beta < 0 \text{ or } \beta > 1 \\ -\log x & \text{if } \beta = 0 \\ 1 - x^\beta & \text{if } 0 < \beta < 1 \\ x \log x & \text{if } \beta = 1 \end{cases}$$

is called *population additively decomposable income inequality index of order* $\beta$.

Theil's inequality index, $\mathbf{I}^1$, is very well-known and used in Economics and Industry, and satisfied that $\mathbf{I}^1(\mathbf{X}^\star) = \lim_{\beta\to 1} \mathbf{I}^\beta(\mathbf{X}^\star)$. On the other hand, $\mathbf{I}^0(\mathbf{X}^\star) = \lim_{\beta\to 0} \mathbf{I}^\beta(\mathbf{X}^\star)$, so that the parametrized family of additive decomposable inequality indices is continuous with respect to the parameter $\beta$.

It should be emphasized that another well-known family of inequality indices, defined by Atkinson [1], and being used for long in Economics, is ordinally equivalent to the family in Definition 2.1.

To introduce additive decomposable indices in [4, 7, 8, 9, 15], and [17], authors first consider a set of desirable properties for inequality measures, having an intuitive interpretation in real-life problems (mainly in those concerning income or wealth inequality, or industrial concentration). In this way, some of those desirable properties are: minimum inequality arises when all individuals in the population have the same attribute value (*normalization*); inequality only depends on ratios of each pair of attribute values (*dimensionless or mean independence*); inequality will decrease if a transfer from a high attribute value to a low one, preserving relation order between the two values, is accomplished (*Pigou–Dalton principle of transfers*); small changes in attribute values entail small changes on inequality (*continuity*); an exchange of attribute values among individuals in the population does not influence inequality (*symmetry*); inequality in a population may be expressed as the sum of inequality among groups determined by a partition of the population, and a kind of average inequality within groups (*additive decomposability*).

Indices in Definition 2.1 can be axiomatically characterized by means of the following properties: normalization, mean independence, continuity, symmetry and additive decomposability, along with some few additional assumptions on the derivativeness of indices with respect to the vector of attribute values.

The structure of the average in the additive decomposability property (which will depend on $\beta$), determines the corresponding additive decomposable index and means the essential difference among indices in the family. On the other hand, the effect of the parameter $\beta$ in that family is that of weighting of the degree of "inequality aversion" (see [1]).

In previous papers ([5, 6, 11] and [12]), we have remarked the importance of the index $\mathbf{I}^{-1}$ in estimating population inequality when only small samples from it are available. However, we are now going to consider situations in which large samples can be drawn from the population.

Assume that a stratified sample of size $n$ is drawn at random from the population independently from different strata. We first suppose that the sample is chosen by a specified allocation $\{\omega_k\},\ k = 1, \dots, r$, so that a sample of size $n_k$ is drawn at random (with or without replacement) from the $k$th stratum, where $n_k/n = \omega_k$. Let $f_{ik}$ and $f_{i.}$ respectively denote the relative frequencies of individuals in the sample from the $k$th stratum and in the sample from the whole population with attribute value equal to $x_i^\star$ $(i = 1, \dots, M,\ k = 1, \dots, r)$.

If $\mathbb{X}^\star$ denotes a random variable whose distribution in the $k$th stratum coincides with that of a random variable $\mathbb{X}_k^\star$ taking on values $x_1^\star, \dots, x_M^\star$ with respective probabilities $f_{1k}, \dots, f_{Mk}$, we have that

**Theorem 2.1.** In the stratified random sampling, the estimator given by

$$I_n^\beta(\mathbb{X}^\star) = \sum_{i=1}^{M}\sum_{k=1}^{r} \frac{W_k}{\omega_k}\, f_{ik}\, \phi_\beta\left(\frac{x_i^\star}{\mathsf{E}(\mathbb{X}^\star)}\right)$$

where

$$\mathsf{E}(\mathbb{X}^\star) = \sum_{i=1}^{M}\sum_{k=1}^{r} \frac{W_k}{\omega_k}\, f_{ik}\, x_i^\star$$

is asymptotically unbiased (as $n_k \to \infty$ for all $k$) to estimate $\mathbf{I}^\beta(\mathbf{X}^\star)$.

In addition, the statistic $n^{\frac{1}{2}}\left[I_n^\beta(\mathbb{X}^\star) - \mathbf{I}^\beta(\mathbf{X}^\star)\right]$ is asymptotically normally distributed (as $n_k \to \infty$, for all $k$) with mean zero and variance equal to

$$(\tau^s)^2 = \sum_{k=1}^{r} \frac{W_k^2}{\omega_k}\left\{\sum_{i=1}^{M} \frac{p_{ik}}{W_k}(V_i)^2 - \left[\sum_{i=1}^{M} \frac{p_{ik}}{W_k}\, V_i\right]^2\right\}$$

where for each $\beta \in \mathbb{R}$,

$$V_i = \phi_\beta\left(\frac{x_i^\star}{\mathsf{E}(\mathbf{X}^\star)}\right) - \frac{x_i^\star}{\mathsf{E}(\mathbf{X}^\star)}\sum_{j=1}^{M} p_{j.}\, \frac{x_j^\star}{\mathsf{E}(\mathbf{X}^\star)}\, \phi'_\beta\left(\frac{x_j^\star}{\mathsf{E}(\mathbf{X}^\star)}\right)$$

($\phi'_\beta$ being the first order derivative function of $\phi_\beta$), whenever $(\tau^s)^2 > 0$.

P r o o f. Indeed, according to some well-known results in large sample theory (see, [2, 3], and [14]), for all $\beta \in \mathbb{R}$, we have that

$$I_n^\beta(\mathbb{X}^\star) - \mathbf{I}^\beta(\mathbf{X}^\star) = \sum_{i=1}^{M}\sum_{k=1}^{r} V_i\left(\frac{W_k}{\omega_k}\, f_{ik} - p_{ik}\right) + R_n$$

where $R_n$ is the corresponding Lagrange remainder term for the first order expansion.

Consequently, the expectation of $I_n^\beta(\mathbb{X}_n) - \mathbf{I}^\beta(\mathbf{X}^*)$ converges to 0 as $n_k \to \infty$ for all $k$, whence $I_n^\beta(\mathbb{X}^\star)$ is asymptotically unbiased to estimate $\mathbf{I}^\beta(\mathbf{X}^\star)$.

On the other hand, under the sampling we have considered, one can guarantee (see, [2, 3], and [14]) that the statistic

$$(n_k)^{\frac{1}{2}} \sum_{i=1}^{M} V_i \left[ \frac{f_{ik}}{\omega_k} - \frac{p_{ik}}{W_k} \right]$$

is asymptotically distributed (as $n_k \to \infty$ for all $k$) as a $N\left(0,\, (T_k)^{\frac{1}{2}}\right)$, where $T_k = {}^tV \sum(k)\, V$ and

$$V = \begin{bmatrix} V_1 \\ \vdots \\ V_k \end{bmatrix}, \qquad \sum(k) = \text{diagonal}\left\{\frac{p_{1k}}{W_k}, \ldots, \frac{p_{Mk}}{W_k}\right\} - P(k)\, {}^tP(k), \qquad P(k) = \begin{bmatrix} \frac{p_{1k}}{W_k} \\ \vdots \\ \frac{p_{Mk}}{W_k} \end{bmatrix}.$$

Therefore, and because of the independence among subsamples from different strata, the asymptotic distribution of the statistic $n^{\frac{1}{2}} \left[ I_n^{\beta}(\mathbb{X}^\star) - \mathbf{I}^{\beta}(\mathbf{X}^\star) \right]$ (as $n_k \to \infty$, for all $k$) is $N(0, \tau^s)$, where

$$(\tau^s)^2 = \sum_{k=1}^{r} \frac{W_k^2}{\omega_k}\, T_k = \sum_{k=1}^{r} \frac{W_k^2}{\omega_k} \left\{ \sum_{i=1}^{M} \frac{p_{ik}}{W_k}\, (V_i)^2 - \left[ \sum_{i=1}^{M} \frac{p_{ik}}{W_k} \right]^2 \right\}$$

whenever $\tau^s > 0$. □

## 3. APPROACHING THE OPTIMUM ALLOCATION

Theorem 2.1 will allow us to asymptotically approach the optimum allocation to estimate $\mathbf{I}^{\beta}(\mathbf{X}^\star)$ on the basis of $I_n^{\beta}(\mathbb{X}^\star)$. Thus, under the assumption that large samples are available, we can get next results.

**Theorem 3.1.** In the stratified random sampling, the asymptotic variance of $I_n^{\beta}(\mathbb{X}^\star)$ is minimized for a fixed total size of sample, $n$, if

$$n_k = \frac{W_k (T_k)^{\frac{1}{2}}}{\sum_{\ell=1}^{r} W_\ell (T_\ell)^{\frac{1}{2}}}\, n$$

where

$$T_k = \left\{ \sum_{i=1}^{M} \frac{p_{ik}}{W_k}\, (V_i)^2 - \left[ \sum_{i=1}^{M} \frac{p_{ik}}{W_k}\, V_i \right]^2 \right\}$$

whenever there is at least one $k \in \{1, \ldots, r\}$ such that $T_k > 0$. The minimum asymptotic variance is then given by

$$\left[ (\tau^s)^2 / n \right]_{\min} = \left[ \sum_{k=1}^{r} W_k (T_k)^{\frac{1}{2}} \right]^2 / \, n.$$

In addition, the whole sample size $n$ is minimized for fixed asymptotic variance $V_0 > 0$ if

$$n_k = \frac{W_k (T_k)^{\frac{1}{2}}}{V_0} \sum_{\ell=1}^{r} W_\ell (T_\ell)^{\frac{1}{2}}$$

and, the minimum total size of the sample is given by

$$n_{\min} = \left[ \sum_{k=1}^{r} W_k (T_k)^{\frac{1}{2}} \right]^2 / \, V_0.$$

P r o o f. Indeed, for all $\beta \in \mathbb{R}$, to minimize $(\tau^s)^2/n$ subject to the constraint $\sum_k n_k = n$, we use Lagrange multipliers method, from which we get the system of equations

$$\begin{cases} \lambda - \frac{W_k^2}{\omega_k^2}\, T_k = 0, \qquad k = 1, \ldots, r \\ \sum\limits_{k=1}^{r} \omega_k = 1 \end{cases}$$

and, hence, the optimum allocation is given by

$$\omega_k = \frac{W_k (T_k)^{\frac{1}{2}}}{\sum\limits_{\ell=1}^{r} W_\ell (T_\ell)^{\frac{1}{2}}}$$

(whenever there is at least one $k \in \{1, \ldots, r\}$ such that $T_k > 0$). The value for the optimum asymptotic variance can be immediately deduced.

In a similar way, to minimize $n$ subject to the constraint $(\tau^s)^2/n = V_0$, we again make use of Lagrange multipliers method, from which we get now the system of equations

$$\begin{cases} 1 - \lambda\, \frac{W_k^2}{n_k^2}\, T_k = 0, \qquad k = 1, \ldots, r \\ \sum\limits_{k=1}^{r} \frac{W_k^2}{n_k}\, T_k = V_0 \end{cases}$$

and, hence, the optimum allocation is given by

$$n_k = \frac{W_k (T_k)^{\frac{1}{2}}}{V_0} \sum_{\ell=1}^{r} W_\ell (T_\ell)^{\frac{1}{2}}$$

(whenever there is at least one $k \in \{1, \ldots, r\}$ such that $T_k > 0$). The value for the optimum whole sample size can be immediately deduced. □

The value $T_k$ in Theorem 3.1 involves unknown population values, that could be replaced by the asymptotically unbiased estimator $\mathbb{T}_k$ defined on the basis of a pilot sample (drawn according to a specified allocation given by $\{\omega_k^0\}$, $k = 1, \ldots, r$, and with relative frequencies $f_{ik}^0$ for the class $x_i$ in the $k$th stratum) by

$$\mathbb{T}_k = \left\{ \sum_{i=1}^{M} \frac{f_{ik}^0}{\omega_k^0} (\mathbb{V}_i^0)^2 - \left[ \sum_{i=1}^{M} \frac{f_{ik}^0}{\omega_k^0} \mathbb{V}_i^0 \right]^2 \right\}$$

where

$$\mathbb{V}_i^0 = \phi_\beta \left( \frac{x_i^\star}{\mathsf{E}(\mathbb{X}^\star)} \right) - \frac{x_i^\star}{\mathsf{E}(\mathbb{X}^\star)} \sum_{j=1}^{M} \sum_{k=1}^{r} f_{ik}^0 \frac{W_k}{\omega_k^0} \frac{x_j^\star}{\mathsf{E}(\mathbb{X}^\star)} \phi'_\beta \left( \frac{x_j^\star}{\mathsf{E}(\mathbb{X}^\star)} \right)$$

and

$$\mathsf{E}(\mathbb{X}^\star) = \sum_{i=1}^{M} \sum_{k=1}^{r} \frac{W_k}{\omega_k^0} f_{ik}^0 x_i^\star .$$

On the other hand, the optimum allocation will be in both cases approached by means of the greatest integer part of the solutions in Theorem 3.1.

It should be emphasized that Theorem 3.1 suggests for both purposes, maximizing precision or minimizing sample size, choosing the sample size in each stratum so that the larger the stratum is, or the higher the stratum variance of the random variable

$\mathbf{V}$ (defined as coinciding in the $k$th stratum with a random variable $\mathbf{V}_k$ taking on values $V_i$ with probability $p_{ik}/W_k,\ i = 1, \ldots, M$), the larger the corresponding "optimum" sample size.

We are finally going to confirm advantages of stratified random sampling with optimum allocation. Thus, if $V_{\text{ran}}$, $V_{\text{prop}}$ and $V_{\text{opt}}$ respectively denote the generic asymptotic variance of $I_n^\beta$ in the simple random sampling, stratified random sampling with proportional allocation, and stratified random sampling with optimum allocation, we then obtain that

**Theorem 3.2.** For fixed total size of the sample, $n$, we have

$$V_{\text{ran}} \geq V_{\text{prop}} \geq V_{\text{opt}}.$$

In addition, $V_{\text{ran}} = V_{\text{prop}}$ iff $\sum_i p_{ik} V_i / W_k$ does not depend on $k$, and $V_{\text{prop}} = V_{\text{opt}}$ iff $T_k$ does not depend on $k$.

Proof. Indeed,

$$V_{\text{ran}} - V_{\text{prop}} = \sum_{k=1}^{r} W_k \left[ \sum_{i=1}^{M} \frac{p_{ik}}{W_k} V_i - \sum_{\ell=1}^{r} \sum_{i=1}^{M} p_{ik} V_i \right]^2$$

and

$$V_{\text{prop}} - V_{\text{opt}} = \sum_{k=1}^{r} W_k \left[ (T_k)^{\frac{1}{2}} - \sum_{\ell=1}^{r} W_\ell \, (T_\ell)^{\frac{1}{2}} \right]^2 .$$

Conditions for equalities $V_{\text{ran}} = V_{\text{prop}}$ and $V_{\text{prop}} = V_{\text{opt}}$, can be easily derived. □

## 4. CONCLUDING REMARKS

A study similar to that in the present paper can be developed in connection with population *diversity* (that can be understood as the population variation with respect to a qualitative attribute, or a quantitative one, the magnitude of whose values is irrelevant for that variation), on the basis of some previous studies (cf. [10]).

On the other hand, when the asymptotic variance in Theorem 2.1 equals zero, we could then extend Zvárová's results [18] and follow ideas in this note.

(Received April 9, 1991.)

REFERENCES

[1] A. B. Atkinson: On the measurement of inequality. J. Econom. Theory *2* (1970), 131–143.

[2] P. Bickel and K. A. Doksum: Mathematical Statistics. Holden–Day, Inc., Oakland 1977.

[3] Y. M. M. Bishop, S. E. Fienberg and P. W. Holland: Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge 1975.

[4] F. Bourguignon: Decomposable income inequality measures. Econometrica *47* (1979), 901–920.

[5] C. Caso and M. A. Gil: Estimating income inequality in the stratified sampling from complete data; Part 1: The unbiased estimation and applications. Kybernetika *25* (1989), 298–311.

[6] C. Caso and M. A. Gil: Estimating income inequality in the stratified sampling from complete data; Part 2: The asymptotic behaviour and the choice of sample size. Kybernetika *25* (1989), 312–319.

[7] F. A. Cowell: On the structure of additive inequality measures. Rev.of Econom.Stud. *47* (1980), 521–531.

[8] F. A. Cowell and K. Kuga: Additivity and the entropy concept. An axiomatic approach to inequality measurement. J. Econom. Theory *25* (1981), 131–143.

[9] W. Eichhorn and W. Gehrig: Measurement of inequality in economics. In: Modern Applied Mathematics – Optimization and Operations Research (B. Korte, ed.), North–Holland, Amsterdam 1982.

[10] M. A. Gil: A note on the stratification and gain in precision in estimating diversity from large samples. Commun. Statist. – Theory and Methods *18* (1989), 1521–1526.

[11] M. A. Gil, C. Caso and P. Gil: Estudio asintótico de una clase de índices de desigualdad muestrales. Trab. de Est. *4* (1989), 95–109.
[12] M. A. Gil, R. Pérez and P. Gil: A family of measures of uncertainty involving utilities: definitions, properties, applications and statistical inferences. Metrika *36* (1989), 129–147.
[13] I. Horowitz: Employment concentration in the common market: an entropy approach. J. Roy. Statist. Soc. Ser. A *133* (1970), 463–679.
[14] C. R. Rao: Linear Statistical Inference and its Applications. Wiley, New York 1973.
[15] A. F. Shorrocks: The class of additively decomposable inequality measures. Econometrica *48* (1980), 613–625.
[16] H. Theil: Economics and Information Theory. North–Holand, Amsterdam 1967.
[17] D. Zagier: On the decomposability of the Gini coefficient and other indices of inequality. Discussion paper No. 108. Projektgruppe Theoretische Modelle. Universität Bonn, 1983.
[18] J. Zvárová: On asymptotic behaviour of a sample estimator of Rnyi's information of order $\alpha$. In: Trans. 6th Prague Conf. on Inf. Theory, Stat. Dec. Func., Rand. Proc., Academia Prague 1973, pp. 919–924.

*Prof. Dr. María Angeles Gil, Dr. Ignacio Martínez, Departamento de Matemticas, Universidad de Oviedo, 33 007 Oviedo. SPAIN.*