KYBERNETIKA — VOLUME 25 (1989), NUMBER 4

ESTIMATING INCOME INEQUALITY IN THE STRATIFIED SAMPLING FROM COMPLETE DATA

Part I. The Unbiased Estimation and Applications*

COVADONGA CASO, MARÍA ANGELES GIL

In order to answer the problem of quantifying the degree of unequalness among the individual shares in the distribution of income, several authors proposed and characterized the well-known additively decomposable measures of relative income inequality. In a previous paper, Pérez, Caso and Gil [12] verified that for a large uncensused population of income earners, the additively decomposable index of order −1 may be readily and unbiasedly estimated from a sample drawn at random from it.

The Part 1 of the present paper deals with the study of the precision of the unbiased estimators in the stratified samplings with and without replacement. Then, some applications are briefly discussed.

The aim of the Second Part is to analyze the asymptotic distribution of the sample income inequality index and to select the suitable sample size to estimate the population index with a specified degree of precision.

1. INTRODUCTION

Several approaches and related indices have been proposed in the literature to measure the inequality in the individual distribution of shares. One of the most commonly accepted approaches is that connecting the Information Theory with the Inequality Measurement, which usually derives inequality indices axiomatically.

In this way, several authors, [2], [3], [4], [6], [14], [16], have considered the problem of defining a minimal number of essential and natural assumptions that a suitable inequality measure should possess and obtaining a family of such measures on the basis of the characterization determined by those assumptions.

In the First Part of the present paper we are going to study the estimation of an index belonging to the family of "additively decomposable measures of income inequality" ([2], [3], [4], [6], [14], [16]) and satisfying a collection of properties we next recall. This estimation will be developed according to a stratified random

* This work was supported in part by the Comisión Asesora de Investigación Científica y Técnica (CAICYT) under the contract PB 85-0401.

sampling in uncensused populations. The interest of stratifying this kind of population is confirmed in the Second Part of this paper, in which we will verify that, when we deal with large samples, the stratification entails a gain in the precision of estimates. In addition, this interest is supported by the fact that many populations to which the estimation of inequality may be applied arise naturally stratified in practice.

The analysis we shall carry out concentrates on obtaining an unbiased estimator, constructed on the basis of the sample income inequality index, and the computation of its exact mean square error. This analysis will be completed with a commentary about some practical relevances.

It is worth emphasizing that we have selected the *additively decomposable index of order* -1, instead of other well-known measures, such as Theil's or Herfindahl's ones, since they do not allow us to define immediately unbiased estimators and to compute its exact precision. In addition, the index of order -1 exhibits properties similar to those for the entropy of degree $\beta = 2$ (or the Gini-Simpson diversity index), [9], that fits appropriately to the estimation problem as it has been pointed out in the literature. (In fact, the index of order -1 can be regarded as a particularization of the directed-divergence of degree $\beta = 2$.)

Nevertheless, the study in the Second Part of this paper could be accomplished, without difficulties, for the other additively decomposable indices.

2. PRELIMINARY CONCEPTS

In this section we are going to recall in a brief way the definition of the index we shall deal with, and the essential and natural properties it satisfies.

Consider a finite uncensused population of N income earners which is divided into r non-overlapping strata. Assume that each individual income for a certain period is positive, $x_1^*, \ldots, x_M^*$ being the possible different income values in the population ($x_i^* > 0$). Let N_k be the number of individuals in the kth stratum (so that, $N_1 + \ldots + N_r = N$) and let p_{ik} and $p_{i.}$ denote the probabilities that a randomly selected individual in the kth stratum and in the whole population, respectively, has an income equal to x_i^* ($i = 1, \ldots, M$, $k = 1, \ldots, r$) in the considered period of time. If we denote the income distribution by X^*, then the income inequality may be quantified as follows:

Definition 2.1. The value $I^{-1}(X^*)$ defined by

$$I^{-1}(X^*) = \sum_{i=1}^{M} \frac{\mathsf{E}(X^*)}{x_i^*} p_{i.} - 1 = \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{r} \sum_{l=1}^{r} \frac{x_j^*}{x_i^*} p_{ik} p_{jl} - 1$$

is called *population additively decomposable income inequality index of order* -1 *concerning* X^*.

Remark 2.1. The terminology in the preceding definition is that adopted by Bourguignon, [2] and Zagier, [16]. According to Cowell, [3], I^{-1} is twice the generalized entropy of order $\beta = -2$ (and, consequently, it is equivalent to the index of order $\varepsilon = 2$, defined by Atkinson, [1]). Following Shorrocks, [14], it is twice the inequality value of degree $c = -1$. Finally, I^{-1} coincides with one of the measures in the broad family in Gastwirth's paper, [7]: the measure associated with the convex function $h(x) = x^{-1}$.

Remark 2.2. The properties below (we have expressed in intuitive terms) are satisfied by the measure in Definition 2.1 and for all additively decomposable indices of income inequality:

Normalization. $I^{-1}(X^*) = 0$, when all individuals have the same income (that is, X^* is degenerate).

Population size independence. $I^{-1}(X^*)$ depends only on the proportions of the population with income in any given range.

Mean independence or income homogeneity. $I^{-1}(X^*)$ depends only on the ratios of the various incomes in X^*.

Pigou-Dalton principle of transfers. A transfer from a rich earner to a poor one, that is not so large as to reverse their positions, should definitively reduce I^{-1}.

Continuity. A small change in the incomes or a change involving only a small proportion of the total population and total income, have a small effect on I^{-1}.

Additive Decomposability. Given a population of income earners and a partition of this population into several non-empty groups, the income inequality in the whole population may be decomposed as the sum of the income inequality "among groups", and a kind of average inequality "within groups". (The formal expression of this property is similar to the recursivity or branching properties satisfied by many information measures.)

This last property (that is not satisfied for some well-known indices of income inequality, such as the Gini measure) is desirable in practice in the inequality measurement, since indices having it not only allow us to compare overall levels of inequality, but also to examine the separate contributions to differences in inequality between two income distributions.

On the basis of the preceding properties, such class of indices is axiomatically characterized (see, for instance, [2], [3], [6], [14] and [16]).

We are next going to estimate unbiasedly the population income inequality index and to compute the precision of the unbiased estimator.

3. UNBIASED ESTIMATION OF THE POPULATION INCOME INEQUALITY. PRECISION OF THE UNBIASED ESTIMATES

Assume that a stratified sample of size n is drawn at random from the population independently in different strata. For the sake of operativeness, we hereafter suppose that the sample is chosen by proportional allocation in each stratum, so that a sample of size n_k is drawn at random (with or without replacement) from the kth stratum, where $n_k/n = N_k/N$, $k = 1, \dots, r$. If f_{ik} and $f_{i.}$ denote the relative frequencies of individuals in the sample from the kth stratum and in the sample from the whole population, respectively, with income equal to x_i^* ($i = 1, \dots, M$, $k = 1, \dots, r$) in the considered period of time, then the income inequality in the sample may be quantified by

Definition 3.1. The analogue estimator defined by

$$I_n^{-1} = \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{r} \sum_{l=1}^{r} \frac{x_j^*}{x_i^*} f_{ik} f_{jl} - 1$$

is called *sample additively decomposable income inequality index of order* -1 *concerning* X^*.

In the same way, we can evaluate the income inequality in the kth stratum as follows:

Definition 3.2. The value $I^{-1}(k)$ defined by

$$I^{-1}(k) = \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{x_j^*}{x_i^*} \left(\frac{Np_{ik}}{N_k}\right) \left(\frac{Np_{jk}}{N_k}\right) - 1$$

is called *additively decomposable income inequality index of order* -1 *concerning* X^* *in the k-th stratum*

Definition 3.3. The analogue estimator defined by

$$I_n^{-1}(k) = \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{x_j^*}{x_i^*} \left(\frac{nf_{ik}}{n_k}\right) \left(\frac{nf_{jk}}{n_k}\right) - 1$$

is called *sample additively decomposable income inequality index of order* -1 *concerning* X^* *in the k-th stratum.*

We are now going to establish exact expressions relating the sample with the population inequality indices (see [12]):

Theorem 3.1. In the stratified random sampling with replacement and proportional allocation, and independently in different strata, the estimator

$$(I_n^{-1})^S = I_n^{-1} + \frac{1}{n^2} \sum_{k=1}^{r} \frac{n_k^2}{n_k - 1} I_n^{-1}(k)$$

is an unbiased estimator of $I^{-1}(X^*)$.

Proof. The expected value of the sample index may be expressed by

$$(1) \qquad \mathrm{E}(I_n^{-1}) = \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{r} \sum_{l=1}^{r} \frac{x_j^*}{x_i^*} \mathrm{E}(f_{ik} f_{jl}) - 1 = \sum_{i=1}^{M} \sum_{k=1}^{r} \mathrm{E}(f_{ik}^2) +$$

$$+ \sum_{i=1}^{M} \sum_{j \neq i} \frac{x_j^*}{x_i^*} \sum_{k=1}^{r} \mathrm{E}(f_{ik} f_{jk}) + \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{x_j^*}{x_i^*} \sum_{k=1}^{r} \sum_{l \neq k} \mathrm{E}(f_{ik} f_{jl}) - 1$$

On the other hand, under the considered sampling, the random vectors $(nf_{1k}, \ldots$ $\ldots, nf_{Mk})$ are independent and follow a multinomial distribution with parameters $n_k, Np_{1k}/N_k, \ldots, Np_{Mk}/N_k$, $(k = 1, \ldots, r)$, and $\mathsf{E}[I_n^{-1}(k)] = (n_k - 1)\, I^{-1}(k)/n_k$.

Finally, if we replace the corresponding moments of the multinomial distribution in (1), the final expression for the unbiased estimator follows. □

Theorem 3.2. In the stratified random sampling without replacement and proportional allocation, and independently in different strata, the estimator

$$(\mathbb{I}_n^{-1})^{SC} = I_n^{-1} + \sum_{k=1}^{r} \frac{n_k^2(N_k - n_k)}{N_k(n_k - 1)}\, I_n^{-1}(k)$$

is an unbiased estimator of $I^{-1}(X^*)$.

Proof. Under the considered sampling, the random vectors $(nf_{1k}, \ldots, nf_{Mk})$ are independent and follow a multivariate hypergeometric distribution with parameters N_k, $D_{1k} = Np_{1k}, \ldots, D_{Mk} = Np_{Mk}$ and n_k, $(k = 1, \ldots, r)$, and $\mathsf{E}[I_n^{-1}(k)] = N_k(n_k - 1)$ $I^{-1}(k)/n_k(N_k - 1)$.

Finally, if we replace the corresponding moments of the multivariate hypergeometric distribution in (1), the final expression for the unbiased estimator follows. □

The knowledge of the distributions associated with the random vectors $(nf_{1k}, \ldots$ $\ldots, nf_{Mk})$ allows us to compute the mean square error of the unbiased estimators in Theorems 3.1 and 3.2. For the sake of simplicity and for practical computational purposes, we have next expressed the mean square errors in terms of values that can be regarded as special cases of some information measures: directed-divergences of degrees $\beta = 2, 3$ regarding two or three special distributions ([11], [13]).

Theorem 3.3. In the stratified random sampling with replacement and proportional allocation, and independently in different strata, the variance of the unbiased estimator $(\mathbb{I}_n^{-1})^S$ is given by

$$\mathsf{Var}[(\mathbb{I}_n^{-1})^S] = \frac{1}{n^4} \sum_{k=1}^{r} \frac{n_k^2}{n_k - 1} \{2[I^{-1}(k)]^2 - 3[I^{-2}(k) + I^{-2}(k^{-1})] + I^{-1}(k^2) +$$

$$+ 2I^{-1}(k)\} + \frac{1}{n^4} \sum_{k=1}^{r} \sum_{l=1}^{r} n_k n_l \{[I^{-1}(k, l)]^2 + I^{-1}(k)\, I^{-1}(l)\} -$$

$$- \frac{1}{n^2} \sum_{k=1}^{r} n_k \{[I^{-1}(k, \cdot)]^2 + [I^{-1}(\cdot, k)]^2 + 2I^{-1}(X^*)\, I^{-1}(k) + 2I^{-1}(k)\} -$$

$$- \frac{3}{n^3} \sum_{k=1}^{r} n_k \{I^{-2}(\cdot, k) + I^{-2}(\cdot, k^{-1})\} + \frac{3}{n} \{I^{-2}(X^*) + I^{-2}(X^{*-1})\} +$$

$$+ \frac{1}{n^2} \{I^{-1}(X^{*2}) - 2(2n - 1)\, I^{-1}(X^*)\}$$

where for $k = 1, \ldots, r,\ l = 1, \ldots, r$,

$$I^{-2}(k) = \frac{1}{3}\left[\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{h=1}^{M}\frac{x_j^* x_h^*}{x_i^{*2}}\left(\frac{Np_{ik}}{N_k}\right)\left(\frac{Np_{jk}}{N_k}\right)\left(\frac{Np_{hk}}{N_k}\right) - 1\right]$$

$$I^{-1}(k,l) = \sum_{i=1}^{M}\sum_{j=1}^{M}\frac{x_j^*}{x_i^*}\left(\frac{Np_{ik}}{N_k}\right)\left(\frac{Np_{jl}}{N_l}\right) - 1$$

$$I^{-1}(k,\cdot) = \sum_{i=1}^{M}\sum_{j=1}^{M}\frac{x_j^*}{x_i^*}\left(\frac{Np_{ik}}{N_k}\right)p_{j.} - 1$$

$$I^{-1}(\cdot, k) = I^{-1}(k^{-1}, \cdot)$$

$$I^{-2}(\cdot, k) = \frac{1}{3}\left[\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{h=1}^{M}\frac{x_j^* x_h^*}{x_i^{*2}}\, p_{i.}\left(\frac{Np_{jk}}{N_k}\right)\left(\frac{Np_{hk}}{N_k}\right) - 1\right]$$

$$I^{-2}(X^*) = \frac{1}{3}\left[\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{h=1}^{M}\frac{x_j^* x_h^*}{x_i^{*2}}\, p_{i.}p_{j.}p_{h.} - 1\right]$$

(X^{*-1}, X^{*2}, k^{-1} and k^2 representing the income distributions taking on the values x_i^{*-1} and x_i^{*2} in the population and the kth stratum, respectively, with the corresponding probabilities in X^* and k.)

Proof. We may express the variance of the unbiased estimator as follows (with $c_k = (n_k - 1)^{-1}$):

$$\mathrm{Var}(\mathbb{I}_n^{-1}) = \mathrm{E}[(I_n^{-1})^2] + \sum_{k=1}^{r}\sum_{l=1}^{r}\left(\frac{n_k}{n}\right)^2\left(\frac{n_l}{n}\right)^2 c_k c_l\,\mathrm{E}[I_n^{-1}(k)\,I_n^{-1}(l)] +$$

$$+ 2\sum_{k=1}^{r}\left(\frac{n_k}{n}\right)^2 c_k\,\mathrm{E}[I_n^{-1}\,I_n^{-1}(k)] - [\mathrm{E}(\mathbb{I}_n^{-1})]^2$$

The briefest possible decomposition of the last member so that the different expressions of the moments about the origin of the multinomial distribution may be immediately replaced is the following:

(2)

$$\mathrm{Var}(\mathbb{I}_n^{-1}) = \sum_{k=1}^{r}\left(\frac{n_k}{n}\right)^4 c_k^2\,\mathrm{E}[(I_n^{-1}(k))^2] + \sum_{k=1}^{r}\sum_{l\neq k}\left(\frac{n_k}{n}\right)^2\left(\frac{n_l}{n}\right)^2 c_k c_l\,\mathrm{E}[I_n^{-1}(k)\,I_n^{-1}(l)] -$$

$$- 2\left[1 + \sum_{k=1}^{r}\left(\frac{n_k}{n}\right)^2 c_k\right]\mathrm{E}(I_n^{-1}) - 2\sum_{k=1}^{r}\left(\frac{n_k}{n}\right)^2 c_k\,\mathrm{E}[I_n^{-1}(k)] - 2\sum_{k=1}^{r}\left(\frac{n_k}{n}\right)^2 c_k - 1 +$$

$$+ \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{h=1}^{M}\sum_{m=1}^{M}\frac{x_h^* x_m^*}{x_i^* x_j^*}\sum_{k=1}^{r}\sum_{l\neq k}\sum_{u\neq k,l}\sum_{v\neq k,l,u}\mathrm{E}(f_{ik}f_{jl}f_{hu}f_{mv}) +$$

$$+ 2\sum_{i=1}^{M}\sum_{j\neq i}\sum_{h=1}^{M}\sum_{m=1}^{M}\frac{x_j^* x_m^*}{x_i^* x_h^*}\sum_{k=1}^{r}\sum_{l\neq k}\sum_{u\neq k,l}(2 + c_k)\,\mathrm{E}(f_{ik}f_{jk}f_{hl}f_{mu}) +$$

$$+\sum_{i=1}^{M}\sum_{j\neq i}\sum_{h=1}^{M}\sum_{m=1}^{M}\left(\frac{x_i^*x_j^*}{x_h^*x_m^*}+\frac{x_h^*x_m^*}{x_i^*x_j^*}\right)\sum_{k=1}^{r}\sum_{l\neq k}\sum_{u\neq k,l}\mathsf{E}(f_{ik}f_{jk}f_{hl}f_{mu})+$$

$$+\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{h=1}^{M}\frac{x_h^*}{x_j^*}\sum_{k=1}^{r}\sum_{l\neq k}\sum_{u\neq k,l}(2+c_k)\,\mathsf{E}(f_{ik}^2f_{jl}f_{hu})+$$

$$+\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{h=1}^{M}\left(\frac{x_j^*x_h^*}{x_i^{*2}}+\frac{x_i^{*2}}{x_j^*x_h^*}\right)\sum_{k=1}^{r}\sum_{l\neq k}\sum_{u\neq k,l}\mathsf{E}(f_{ik}^2f_{jl}f_{hu})+$$

$$+\sum_{i=1}^{M}\sum_{j\neq i}\sum_{h=1}^{M}\sum_{m\neq h}\frac{x_h^*x_m^*}{x_i^*x_j^*}\sum_{k=1}^{r}\sum_{l\neq k}\mathsf{E}(f_{ik}f_{jk}f_{hl}f_{ml})+$$

$$+\,2\sum_{i=1}^{M}\sum_{j\neq i}\sum_{h=1}^{M}\sum_{m\neq h}\frac{x_j^*x_m^*}{x_i^*x_h^*}\sum_{k=1}^{r}\sum_{l\neq k}(1+c_k)\,\mathsf{E}(f_{ik}f_{jk}f_{hl}f_{ml})+$$

$$+\,2\sum_{i=1}^{M}\sum_{j\neq i}\sum_{h\neq i,j}\sum_{m=1}^{M}\left(\frac{x_h^*x_m^*}{x_i^*x_j^*}+\frac{x_i^*x_h^*}{x_j^*x_m^*}\right)\sum_{k=1}^{r}\sum_{l\neq k}(1+c_k)\,\mathsf{E}(f_{ik}f_{jk}f_{hk}f_{ml})+$$

$$+\sum_{i=1}^{M}\sum_{j\neq i}\sum_{h=1}^{M}\left(\frac{2x_j^*}{x_h^*}+\frac{2x_h^*}{x_j^*}+\frac{x_j^*x_h^*}{x_i^{*2}}+\frac{x_i^{*2}}{x_j^*x_h^*}\right)\sum_{k=1}^{r}\sum_{l\neq k}(1+c_k)\,\mathsf{E}(f_{ik}^2f_{jk}f_{hl})+$$

$$+\,2\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{h\neq j}\frac{x_h^*}{x_j^*}\sum_{k=1}^{r}\sum_{l\neq k}(2+c_k+c_l)\,\mathsf{E}(f_{ik}^2f_{jl}f_{hl})+$$

$$+\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{h\neq j}\left(\frac{x_j^*x_h^*}{x_i^{*2}}+\frac{x_i^{*2}}{x_j^*x_h^*}\right)\sum_{k=1}^{r}\sum_{l\neq k}\mathsf{E}(f_{ik}^2f_{jl}f_{hl})+$$

$$+\,2\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\frac{x_j^*}{x_i^*}+\frac{x_i^*}{x_j^*}\right)\sum_{k=1}^{r}\sum_{l\neq k}(1+c_k)\,\mathsf{E}(f_{ik}^3f_{jl})+$$

$$+\,2\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{r}\sum_{l\neq k}(1+c_k)\,\mathsf{E}(f_{ik}^2f_{jl}^2)+\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{x_j^{*2}}{x_i^{*2}}\sum_{k=1}^{r}\sum_{l\neq k}\mathsf{E}(f_{ik}^2f_{jl}^2)+$$

$$+\sum_{i=1}^{M}\sum_{j\neq i}\sum_{h\neq i,j}\sum_{m\neq i,j,h}\frac{x_h^*x_m^*}{x_i^*x_j^*}\sum_{k=1}^{r}(1+2c_k)\,\mathsf{E}(f_{ik}f_{jk}f_{hk}f_{mk})+$$

$$+\sum_{i=1}^{M}\sum_{j\neq i}\sum_{h\neq i,j}\left(\frac{4x_h^*}{x_j^*}+\frac{x_j^*x_h^*}{x_i^{*2}}+\frac{x_i^{*2}}{x_j^*x_h^*}\right)\sum_{k=1}^{r}(1+2c_k)\,\mathsf{E}(f_{ik}^2f_{jk}f_{hk})+$$

$$+\,2\sum_{i=1}^{M}\sum_{j\neq i}\left(\frac{x_j^*}{x_i^*}+\frac{x_i^*}{x_j^*}\right)\sum_{k=1}^{r}(1+2c_k)\cdot\mathsf{E}(f_{ik}^3f_{jk})+$$

$$+\sum_{i=1}^{M}\sum_{j\neq i}\left(2+\frac{x_j^{*2}}{x_i^{*2}}\right)\sum_{k=1}^{r}(1+2c_k)\,\mathsf{E}(f_{ik}^2f_{jk}^2)+$$

$$+\sum_{i=1}^{M}\sum_{k=1}^{r}(1+2c_k)\,\mathsf{E}(f_{ik}^4)-[I^{-1}(X^*)]^2.$$

On the basis of this expression and considering the corresponding moments of the multinomial distribution we obtain the final formula for the mean square error of $(\mathbb{I}_n^{-1})^S$. □

(Obviously, $\boldsymbol{I}^{-1}(k)$ and $\boldsymbol{I}^{-1}(X^*)$ may be regarded as particularizations of the directed-divergence of degree $\beta = 2$ between two distributions, $\boldsymbol{I}^{-1}(k, l)$, $\boldsymbol{I}^{-1}(k, \cdot)$ and $\boldsymbol{I}^{-1}(\cdot, k)$ may be regarded as particularizations of the directed-divergence of degree $\beta = 2$ among three distributions, and $\boldsymbol{I}^{-2}(\cdot, k)$ may be regarded as a particularization of the directed-divergence of degree $\beta = 3$ among three distributions.)

Theorem 3.4. In the stratified random sampling without replacement and proportional allocation, and independently in different strata, the variance of the unbiased estimator $(\mathbb{I}_n^{-1})^{SC}$ is given by

$$V_n^C = \mathrm{Var}[(\mathbb{I}_n^{-1})^{SC}] = \frac{1}{n^4} \sum_{k=1}^{r} \frac{n_k^2(N_k - n_k)}{(n_k - 1) N_k(N_k - 1)^2 (N_k - 2)(N_k - 3)} \cdot$$

$$\cdot \{2[N_k^4 + (3n_k - 5) N_k^3 - (7n_k^2 - 2n_k - 6) N_k^2 + 9n_k(n_k - 1) N_k] [\boldsymbol{I}^{-1}(k)]^2 -$$

$$- 3[N_k^4 + (2n_k - 5) N_k^3 - (5n_k^2 - 2n_k - 6) N_k^2 +$$

$$+ n_k(7n_k - 8) N_k] [\boldsymbol{I}^{-2}(k) + \boldsymbol{I}^{-2}(k^{-1})] + [N_k^4 - 5N_k^3 - (2n_k^2 - 5n_k - 6)N_k^2 +$$

$$+ n_k(3n_k - 10) N_k + n_k(n_k + 1)] \boldsymbol{I}^{-1}(k^2) +$$

$$+ 2[N_k^4 + (6n_k - 5) N_k^3 - (11n_k^2 + 4n_k - 6) N_k^2 +$$

$$+ n_k(15n_k - 4) N_k - 2n_k(n_k + 1)] \boldsymbol{I}^{-1}(k)\} +$$

$$+ \frac{1}{n^4} \sum_{k=1}^{r} \sum_{l=1}^{r} n_k n_l \frac{(N_k - n_k)(N_l - n_l)}{(N_k - 1)(N_l - 1)} \cdot$$

$$\cdot \{[\boldsymbol{I}^{-1}(k, l)]^2 + \boldsymbol{I}^{-1}(k) \,.\, \boldsymbol{I}^{-1}(l) - 3[\boldsymbol{I}^{-2}(k, l) + \boldsymbol{I}^{-2}(k^{-1}, l^{-1})] +$$

$$+ \boldsymbol{I}^{-1}(k^2, l^2) + 2\boldsymbol{I}^{-1}(k, l)\} -$$

$$- \frac{1}{n^2} \sum_{k=1}^{r} \frac{n_k(N_k - n_k)}{(N_k - 1)} \{2\boldsymbol{I}^{-1}(X^*) \,.\, \boldsymbol{I}^{-1}(k) + [\boldsymbol{I}^{-1}(k, \cdot)]^2 +$$

$$+ [\boldsymbol{I}^{-1}(\cdot, k)]^2 - 3[\boldsymbol{I}^{-2}(k, \cdot) + \boldsymbol{I}^{-2}(k^{-1}, \cdot)] + 2[\boldsymbol{I}^{-1}(k, \cdot) + \boldsymbol{I}^{-1}(\cdot, k) + \boldsymbol{I}^{-1}(k)]\} -$$

$$- \frac{2}{n^3} \sum_{k=1}^{r} \frac{n_k^2(N_k - n_k)}{(N_k - 1)(N_k - 2)} \{2\boldsymbol{I}^{-1}(k) \,.\, [\boldsymbol{I}^{-1}(k, \cdot) + \boldsymbol{I}^{-1}(\cdot, k)] - [1 + \boldsymbol{I}^{-1}(k)] \,.$$

$$. [\boldsymbol{I}^{-1}(\cdot, k^{-1}, k^{-2}) + \boldsymbol{I}^{-1}(\cdot, k, k^2)] + \boldsymbol{I}^{-1}(k, \cdot) + \boldsymbol{I}^{-1}(\cdot, k) + 2\boldsymbol{I}^{-1}(k)\}$$

where for $k = 1, \ldots, r$, $l = 1, \ldots, r$,

$$\boldsymbol{I}^{-2}(k, l) = \frac{1}{3}\left[\sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{h=1}^{M} \frac{x_j^* x_h^*}{x_i^{*2}} \left(\frac{Np_{ik}}{N_k}\right)\left(\frac{Np_{jl}}{N_l}\right)\left(\frac{Np_{hl}}{N_l}\right) - 1\right]$$

$$\boldsymbol{I}^{-2}(k, \cdot) = \frac{1}{3}\left[\sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{h=1}^{M} \frac{x_j^* x_h^*}{x_i^{*2}} \left(\frac{Np_{ik}}{N_k}\right) p_{j.} p_{h.} - 1\right]$$

$$I^{-1}(\cdot, k, k^2) = \frac{1}{1 + I^{-1}(k)} \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{h=1}^{M} \frac{x_i^{*2}}{x_j^* x_h^*} \left(\frac{Np_{ik}}{N_k}\right)\left(\frac{Np_{jk}}{N_k}\right) p_{h.} - 1$$

$$I^{-1}(\cdot, k^{-1}, k^{-2}) = \frac{1}{1 + I^{-1}(k)} \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{h=1}^{M} \frac{x_j^* x_h^*}{x_i^{*2}} \left(\frac{Np_{ik}}{N_k}\right)\left(\frac{Np_{jk}}{N_k}\right) p_{h.} - 1$$

($\boldsymbol{k}^{-2}$ representing the income distribution taking on the values x_i^{*-2} in the kth stratum, with the corresponding probabilities in $\boldsymbol{k}$).

Proof. The final expression for the variance of $(\mathbb{I}_n^{-1})^{SC}$ is obtained by replacing the corresponding moments of the multivariate hypergeometric distribution in (2) with $c_k = (N_k - n_k)/[N_k(n_k - 1)]$. □

(Obviously, $I^{-1}(\cdot, \boldsymbol{k}, \boldsymbol{k}^2)$ and $I^{-1}(\cdot, \boldsymbol{k}^{-1}, \boldsymbol{k}^{-2})$ may be regarded as particularizations of the directed-divergence of degree $\beta = 2$ among three distributions, and $I^{-2}(\boldsymbol{k}, \boldsymbol{l})$, $I^{-2}(\boldsymbol{k}, \cdot)$ may be regarded as particularizations of the directed-divergence of degree $\beta = 3$ among three distributions.)

In practice, the interest of quantifying the precision of the preceding unbiased estimators is due to the fact that their knowledge enables us to compare the precision of these estimates against other ones, to compare the precision of the adopted sampling procedures or allocations, and to choose the suitable sample size to estimate the income inequality with a specified degree of precision.

As the expressions in Theorems 3.3. and 3.4 involve unknown population values, it would be interesting to approximate them by means of analogue estimates usually based on a previous sample drawn at random from the same or a similar population according to a stratified sampling. The Appendix in the First Part of the present paper collects such unbiased estimates.

4. APPLICATIONS AND CONCLUDING COMMENTS

The available information concerning the income distribution in a sample is often supplied in practice according to grouped data. The studies about the estimation of the population income inequality is then carried out ([5], [7], [8]) by establishing upper and lower bounds for income inequality in each interval.

The study in this paper has been developed in terms of the estimation of income inequality of a large uncensused population of income earners, on the basis of complete information. Nevertheless, the results we have obtained may be immediately applied to estimate the *industrial concentration* ([10], [15]). In such a situation, we consider a certain industry with N firms, where the values x_i^* in the preceding sections may be given by the number of employees per firm, the (physical or financial) output, and so on.

APPENDIX 1: APPROXIMATING THE POPULATION VALUES IN THE SQUARE ERROR OF ESTIMATES (STRATIFIED SAMPLING WITH REPLACEMENT)

In the stratified random sampling with replacement and proportional allocation, we have that

$$\frac{n_k}{(n_k - 1)(n_k - 2)(n_k - 3)} \{n_k^2[I_n^{-1}(k)]^2 - 3n_k[I_n^{-2}(k) + I_n^{-2}(k^{-1})] + + 2(3n_k - 2)I_n^{-1}(k) + I_n^{-1}(k^2)\}$$

is an unbiased estimator of $[\boldsymbol{I}^{-1}(\boldsymbol{k})]^2$;

$$\frac{n_k}{3(n_k - 1)(n_k - 2)} \{3n_k I_n^{-2}(k) - 2I_n^{-1}(k) - I_n^{-1}(k^2)\}$$

is an unbiased estimator of $\boldsymbol{I}^{-2}(\boldsymbol{k})$;

$$\frac{n_k}{n_k - 1} I_n^{-1}(k)$$

is an unbiased estimator of $\boldsymbol{I}^{-1}(\boldsymbol{k})$;

$$\frac{1}{(n_k - 1)(n_l - 1)} \{n_k n_l[I_n^{-1}(k, l)]^2 - 3n_k I_n^{-2}(l^{-1}, k^{-1}) - 3n_l I_n^{-2}(k, l) + I_n^{-1}(k^2, l^2) + + 2(n_k + n_l - 1) I_n^{-1}(k, l)\}$$

is an unbiased estimator of $[\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l})]^2$, when $l \neq k$;

$$\frac{n_k n_l}{(n_k - 1)(n_l - 1)} I_n^{-1}(k) I_n^{-1}(l)$$

is an unbiased estimator of $\boldsymbol{I}^{-1}(\boldsymbol{k}) . \boldsymbol{I}^{-1}(\boldsymbol{l})$, when $l \neq k$;

$$\frac{1}{n_k - 1} \{n_k[I_n^{-1}(k, \cdot)]^2 - 3 I_n^{-2}(k, \cdot) + 2 I_n^{-1}(k, \cdot)\} + \frac{1}{n(n_k - 1)} \cdot$$
$$\cdot \{I_n^{-1}(k^2, \cdot) - 3n_k I_n^{-2}(\cdot, k^{-1})\} + \frac{2n_k^2}{n^2(n_k - 1)^2 (n_k - 2)(n_k - 3)} \cdot$$
$$\cdot \{(5n_k - 6)[n_k[I_n^{-1}(k)]^2 - 3 I_n^{-2}(k)] - (2n_k - 3)[3n_k I_n^{-2}(k^{-1}) + I_n^{-1}(k^2)] +$$
$$+ 2(3n_k^2 - n_k - 3) I_n^{-1}(k)\} + \frac{2n_k}{n(n_k - 1)(n_k - 2)} \cdot$$
$$\cdot \{2n_k I_n^{-1}(k) I_n^{-1}(k, \cdot) - 2[1 + I^{-1}(k)] I^{-1}(\cdot, k^{-1}, k^{-2}) + 2 I_n^{-1}(k, \cdot) + n_k I_n^{-1}(k)\} +$$
$$+ \frac{1}{n^2(n_k - 1)} \sum_{l=1}^{r} \frac{n_l}{n_l - 1} \{n_k n_l[I_n^{-1}(k, l)]^2 - 3n_k I_n^{-2}(l^{-1}, k^{-1}) - 3n_l I_n^{-2}(k, l) +$$
$$+ 2n_k n_l I_n^{-1}(k, l) + I_n^{-1}(k^2, l^2)\}$$

is an unbiased estimator of $[I^{-1}(k,\cdot)]^2$;

$$\frac{n_k}{n_k-1} I_n^{-1}(k) I_n^{-1} + \frac{2n_k^3(5n_k-6)}{n^2(n_k-1)^2(n_k-2)(n_k-3)} [I_n^{-1}(k)]^2 -$$

$$- \frac{n_k^3}{n^2(n_k-1)(n_k-2)(n_k-3)} \{9 I_n^{-2}(k) + 9 I_n^{-2}(k^{-1}) - I_n^{-1}(k^2) - 14 I_n^{-1}(k)\} +$$

$$+ \frac{n_k^2}{n(n_k-1)(n_k-2)} \{2 I_n^{-1}(k) [I_n^{-1}(k,\cdot) + I_n^{-1}(\cdot,k)] - [1 + I_n^{-1}(k)] \cdot$$

$$\cdot [I_n^{-1}(\cdot,k^{-1},k^{-2}) + I_n^{-1}(\cdot,k,k^2)] + 2 I_n^{-1}(k) + I_n^{-1}(k,\cdot) + I_n^{-1}(\cdot,k)\} +$$

$$+ \frac{n_k}{n^2(n_k-1)} \sum_{l=1}^{r} \frac{n_l^2}{n_l-1} I_n^{-1}(k) I_n^{-1}(l)$$

is an unbiased estimator of $I^{-1}(k) \,.\, I^{-1}(X^*)$;

$$\frac{1}{3(n_k-1)} \{3n_k I_n^{-2}(\cdot,k) - I_n^{-1}(\cdot,k^2)\} + \frac{2n_k}{3n(n_k-1)(n_k-2)} \cdot$$

$$\cdot \{3n_k I_n^{-2}(k) - I_n^{-1}(k^2) - n_k I_n^{-1}(k)\}$$

is an unbiased estimator of $I^{-2}(\cdot,k)$;

$$I_n^{-2} - \frac{1}{3n}\{2 I_n^{-1} + I_n^{-1}(X^{*2})\} + \frac{1}{3n^2} \sum_{k=1}^{r} \frac{n_k}{n_k-1} \{2n_k[1 + I_n^{-1}(k)] I_n^{-1}(\cdot,k^{-1},k^{-2}) +$$

$$+ 3n_k I_n^{-2}(\cdot,k) - 2 I_n^{-1}(k,\cdot) - I_n^{-1}(\cdot,k^2) + 2n_k I_n^{-1}(k)\} +$$

$$+ \frac{2}{3n^3} \sum_{k=1}^{r} \frac{n_k^3}{(n_k-1)(n_k-2)} \{6 I_n^{-2}(k) - 2 I_n^{-1}(k) - I_n^{-1}(k^2)\}$$

is an unbiased estimator of $I^{-2}(X^*)$;

$$I_n^{-1} + \frac{1}{n^2} \sum_{k=1}^{r} \frac{n_k^2}{n_k-1} I_n^{-1}(k)$$

is an unbiased estimator of $I^{-1}(X^*)$.

APPENDIX 2: APPROXIMATING THE POPULATION VALUES IN THE SQUARE ERROR OF ESTIMATES (STRATIFIED SAMPLING WITHOUT REPLACEMENT)

In the stratified random sampling without replacement and proportional allocation, we have that

$$\frac{n_k(N_k-1)}{N_k^3(n_k-1)(n_k-2)(n_k-3)} \{n_k^2(N_k-2)(N_k-3) [I_n^{-1}(k)]^2 -$$

$$- 3n_k(N_k-n_k)(N_k-2) [I_n^{-2}(k) + I_n^{-2}(k^{-1})] +$$

$$+ 2(N_k - n_k)\left[(3n_k - 2)N_k - 2(2n_k + 1)\right] .$$
$$. I_n^{-1}(k) + (N_k - n_k)(N_k - n_k + 1) I_n^{-1}(k^2)\}$$

is an unbiased estimator of $[\boldsymbol{I}^{-1}(\boldsymbol{k})]^2$;

$$\frac{n_k(N_k - 1)}{3N_k^2(n_k - 1)(n_k - 2)} \{3n_k(N_k - 2) I_n^{-2}(k) - 2(N_k - n_k) I_n^{-1}(k) - (N_k - n_k) I_n^{-1}(k^2)\}$$

is an unbiased estimator of $\boldsymbol{I}^{-2}(\boldsymbol{k})$;

$$\frac{n_k(N_k - 1)}{N_k(n_k - 1)} I_n^{-1}(k)$$

is an unbiased estimator of $\boldsymbol{I}^{-1}(\boldsymbol{k})$;

$$\frac{1}{N_k N_l(n_k - 1)(n_l - 1)} \{n_k n_l(N_k - 1)(N_l - 1)[I_n^{-1}(k, l)]^2 - 3n_l(N_k - n_k)(N_l - 1) I_n^{-2}(k, l) -$$
$$- 3n_k(N_l - n_l)(N_k - 1) I_n^{-2}(l^{-1}, k^{-1}) + (N_k - n_k)(N_l - n_l) I_n^{-1}(k^2, l^2) -$$
$$- 2[N_k N_l(n_k - 1)(n_l - 1) - n_k n_l(N_k - 1)(N_l - 1)] I_n^{-1}(k, l)\}$$

is an unbiased estimator of $[\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l})]^2$, when $l \neq k$;

$$\frac{n_k n_l(N_k - 1)(N_l - 1)}{N_k N_l(n_k - 1)(n_l - 1)} I_n^{-1}(k) I_n^{-1}(l)$$

is an unbiased estimator of $\boldsymbol{I}^{-1}(\boldsymbol{k}) . \boldsymbol{I}^{-1}(\boldsymbol{l})$, when $l \neq k$;

$$\frac{1}{3N_l(n_l - 1)} \{3n_l(N_l - 1) I_n^{-2}(k, l) - (N_l - n_l) I_n^{-1}(k^2, l^2)\}$$

is an unbiased estimator of $\boldsymbol{I}^{-2}(\boldsymbol{k}, \boldsymbol{l})$, when $l \neq k$;

$$I_n^{-1}(k, l)$$

is an unbiased estimator of $\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l})$, when $l \neq k$;

$$\frac{n_k(N_k - 1)}{N_k(n_k - 1)} I_n^{-1}(k) I_n^{-1} + \frac{2n_k^3(N_k - n_k)(N_k - 1)[(5n_k - 6)N_k - 3n_k(n_k - 1)]}{n^2 N_k^3(n_k - 1)^2(n_k - 2)(n_k - 3)} .$$
$$. [I_n^{-1}(k)]^2 - \frac{n_k^3(N_k - n_k)(N_k - 1)}{n^2 N_k^3(n_k - 1)(n_k - 2)(n_k - 3)} \{3(3N_k - 2n_k)[I_n^{-2}(k) +$$
$$+ I_n^{-2}(k^{-1})] - (N_k - n_k + 1) I_n^{-1}(k^2) - 2[7N_k - 2(2n_k + 1)] I_n^{-1}(k)\} +$$
$$+ \frac{n_k^2(N_k - n_k)(N_k - 1)}{nN_k^2(n_k - 1)(n_k - 2)} \{2 I_n^{-1}(k)[I_n^{-1}(k, \cdot) + I_n^{-1}(\cdot, k)] -$$
$$- [1 + I_n^{-1}(k)] . [I_n^{-1}(\cdot, k^{-1}, k^{-2}) + I_n^{-1}(\cdot, k, k^2)] +$$
$$+ I_n^{-1}(k, \cdot) + I_n^{-1}(\cdot, k) + 2 I_n^{-1}(k)\} +$$
$$+ \frac{n_k(N_k - 1)}{n^2 N_k(n_k - 1)} \sum_{l=1}^{r} \frac{n_l^2(N_l - n_l)}{N_l(n_l - 1)} I_n^{-1}(k) I_n^{-1}(l)$$

The unbiased estimators we have just described are expressed in terms of the analogue estimates of the directed-divergence measures of degrees $\beta = 2, 3$ defined in Section 3.

(Received March 29, 1988.)

REFERENCES

[1] A. B. Atkinson: On the measurement of inequality. J. Econom. Theory *2* (1970), 244—263.

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

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

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

[5] F. A. Cowell and F. Mehta: The estimation and interpolation of inequality measures. Rev. Econom. Stud. *49* (1982), 273—290.

[6] 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.

[7] J. L. Gastwirth: The estimation of a family of measures of income inequality. J. Econometrics *3* (1975), 61—70.

[8] J. L. Gastwirth, T. K. Nayak and A. M. Krieger: Large sample theory for the bounds on the Gini and related indices of inequality estimated from grouped data. J. Buss. & Econ. Stat. *4* (1986), 2, 269—273.

[9] J. Havrda and F. Charvát: Quantification method of classification processes. Kybernetika *3* (1967), 30—35.

[10] I. Horowitz: Employment concentration in the Common Market: an entropy approach. J. Roy. Statist. Soc. Ser. A. *133* (1970), 463—479.

[11] A. M. Mathai and P. N. Rathie: Basic Concepts in Information Theory and Statistics. Wiley Eastern Limited, New Delhi 1975.

[12] R. Pérez, C. Caso and M. A. Gil: Unbiased estimation of income inequality. Statist. Hefte *27* (1986), 227—237.

[13] P. N. Rathie and Pl. Kannappan: A directed divergence function of type β. Inform. and Control *20* (1972), 38—45.

[14] A. F. Shorrocks: The class of additively decomposable inequality measures. Econometrica *48* (1980), 613—625.

[15] H. Theil: Economics and Information Theory. North-Holland, Amsterdam 1967.

[16] D. Zagier: On the decomposability of the Gini coefficient and other indices of inequality. Discussion paper No. 108. Projektgruppe Theoretische Modelle. Universität Bonn, 1983.

Dr. Covadonga Caso, Dr. María Angeles Gil, Departamento de Matemáticas, Universidad de Oviedo, 33071 Oviedo. Spain.