Covadonga Caso; María Angeles Gil

Estimating income inequality in the stratified sampling from complete data. II. The asymptotic behaviour and the choice of sample size

Kybernetika, Vol. 25 (1989), No. 4, 312--319

Persistent URL: http://dml.cz/dmlcz/125822

ESTIMATING INCOME INEQUALITY IN THE STRATIFIED SAMPLING FROM COMPLETE DATA

Part II. The Asymptotic Behaviour and the Choice of Sample Size*

COVADONGA CASO, MARÍA ANGELES GIL

The First Part of this paper deals with the study of the precision of an unbiased estimator of a population income inequality index in the stratified samplings with and without replacement.

In this Second Part, we first analyze the asymptotic distribution of the sample income inequality index. Then, we establish different criteria to select the suitable sample size to estimate the population index with a desired degree of precision. One of these criteria is based on Chebyshev's approach and the other one follows from the preceding asymptotic study.

1. INTRODUCTION

When we try to draw statistical conclusions regarding the income inequality in an uncensused population, it would be interesting to know the distribution of the sample index for different sampling methods. Nevertheless, such a distribution cannot be exactly determined, so that the study of the asymptotic behaviour of the sample index, we are going to develop in the next section for the stratified sampling, may be useful. In addition, in most of the problems that try to obtain conclusions about the population income inequality, large samples are available, and the suitability of the preceding study is then justified.

Once we have examined the asymptotic distribution of the sample index, we may readily derive some practical results.

In this way, we are first going to develop a non-conservative procedure, which allows us to select the sample size needed to estimate the population income inequality with a specified degree of precision. This procedure is then compared with the conservative criterion based on Chebyshev's approach.

Other statistical procedures, such as confidence intervals and hypotheses testing methods about the population inequality, are finally suggested.

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

2. ASYMPTOTIC BEHAVIOUR OF THE SAMPLE INCOME INEQUALITY INDEX

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.

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 $n_k/n = N_k/N$, $k = 1, \ldots, r$. Let 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^* $(i = 1, \ldots, M,\ k = 1, \ldots, r)$ in the considered period of time.

According to Definitions 2.1 and 3.1 in the First Part of this paper, [2], $\boldsymbol{I}^{-1}(\boldsymbol{X}^*)$ and I_n^{-1} represent, respectively, the population and sample additively decomposable income inequality index of order -1 (cf., [1], [3], [4], [5], [13], [14]). Following ideas in [11], [12], and [15], we can now establish

Theorem 2.1. The random variable $n^{1/2}[I_n^{-1} - \boldsymbol{I}^{-1}(\boldsymbol{X}^*)]$ is asymptotically distributed (as $n_k \to \infty$, $k = 1, \ldots, r$) according to a normal distribution with mean zero and variance equal to

$$\tau^2 = -\sum_{k=1}^{r} \frac{N_k}{N} \{[\boldsymbol{I}^{-1}(k, \cdot)]^2 + [\boldsymbol{I}^{-1}(\cdot, k)]^2 + 2[\boldsymbol{I}^{-1}(\boldsymbol{X}^*) + 1]\, \boldsymbol{I}^{-1}(k)\} +$$
$$+ 3[\boldsymbol{I}^{-2}(\boldsymbol{X}^*) + \boldsymbol{I}^{-2}(\boldsymbol{X}^{*-1})] - 4\,\boldsymbol{I}^{-1}(\boldsymbol{X}^*)$$

whenever $\tau^2 > 0$.

Proof. The first order Taylor expansion for $I^{-1}(\boldsymbol{f}) \equiv I_n^{-1}(\boldsymbol{f} \equiv (f_{11}, \ldots, f_{M1}, \ldots \ldots, f_{1r}, \ldots, f_{Mr}))$ in a neighborhood of $\boldsymbol{p} \equiv (p_{11}, \ldots, p_{M1}, \ldots, p_{1r}, \ldots, p_{Mr})$ is given by

$$I^{-1}(\boldsymbol{f}) = I^{-1}(\boldsymbol{p}) + \sum_{i=1}^{M-1} \sum_{k=1}^{r} \frac{\partial B(\boldsymbol{p}^*)}{\partial p_{ik}} (f_{ik} - p_{ik}) + R_n$$

where $\boldsymbol{p}^*$ is the $(M-1) \times r$-dimensional vector $(p_{11}, \ldots, p_{(M-1)1}, \ldots, p_{1r}, \ldots, p_{(M-1)r})$, $B(\boldsymbol{p}^*) \equiv I^{-1}(\boldsymbol{p}) \equiv \boldsymbol{I}^{-1}(\boldsymbol{X}^*)$ and R_n is the Lagrange remainder term.

Thus,

$$I^{-1}(\boldsymbol{f}) = I^{-1}(\boldsymbol{p}) + \sum_{i=1}^{M} v_i(f_{i.} - p_{i.}) + R_n, \quad \text{where} \quad v_i = \sum_{j=1}^{M} p_{j.}\left(\frac{x_j^*}{x_i^*} + \frac{x_i^*}{x_j^*}\right)$$

As in the sampling we have considered the random vectors $(n^{1/2}f_{11}, \ldots, n^{1/2}f_{M1}), \ldots$ $\ldots, (n^{1/2}f_{1r}, \ldots, n^{1/2}f_{Mr})$ are independent and each of them has an asymptotic multivariate normal distribution with mean vectors $(n^{1/2}p_{11}, \ldots, n^{1/2}p_{M1}), \ldots, (n^{1/2}p_{1r}, \ldots$ $\ldots, n^{1/2}p_{Mr})$, respectively, then the random vector $(n^{1/2}f_{1.}, \ldots, n^{1/2}f_{M.})$ has an asymptotic multivariate normal distribution with mean vector $(n^{1/2}p_{1.}, \ldots, n^{1/2}p_{M.})$, as a consequence of the reproductivity of the multivariate normal distribution.

The remainder term $n^{1/2}R_n$ can be expressed by

$$n^{1/2}R_n = \sum_{j=2}^{\infty} \sum_{i=1}^{M-1} \sum_{k=1}^{r} \frac{\partial^{j)}B(p^*)}{\partial p_{ik}} n^{1/2}(f_{ik} - p_{ik})^j =$$

$$= \sum_{j=2}^{\infty} \sum_{i=1}^{M-1} \sum_{k=1}^{r} \frac{\partial^{j)}B(p^*)}{\partial p_{ik}} [p_{ik}(1 - p_{ik})]^{1/2} \frac{nf_{ik} - np_{ik}}{[np_{ik}(1 - p_{ik})]^{1/2}} (f_{ik} - p_{ik})^{j-1}$$

As $n \to \infty$, $(f_{ik} - p_{ik})^{j-1}$ converges in probability to zero (weak laws of large numbers) and $[nf_{ik} - np_{ik}]/[np_{ik}(1 - p_{ik})]^{1/2}$ converges in law to a standard normal random variable (De Moivre's Theorem). Then, according to the well-known properties of the convergence in probability we have that $n^{1/2}(f_{ik} - p_{ik})^j$ converges in probability to zero as $n \to \infty$, for all $j = 2, 3, \ldots, i = 1, \ldots, M - 1, k = 1, \ldots, r$, so that $n^{1/2}R_n$ converges in probability to zero as $n \to \infty$.

Therefore, the random variable $n^{1/2}[I_n^{-1} - I^{-1}(X^*)]$, coinciding with $\sum_i n^{1/2}v_i \,.$ $.\,(f_{i.} - p_{i.})$ unless for the remainder term $n^{1/2}R_n$, will be asymptotically distributed according to a normal distribution with mean zero and variance equal to

$$\tau^2 = n\sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{r} \sum_{l=1}^{r} v_i v_j \, \mathsf{E}[(f_{ik} - p_{ik})(f_{jl} - p_{jl})] =$$

$$= -\sum_{k=1}^{r} \frac{N_k}{N} [2 + I^{-1}(k, \cdot) + I^{-1}(\cdot, k)]^2 + 3[I^{-2}(X^*) + I^{-2}(X^{*-1})] + 2I^{-1}(X^*) + 4 =$$

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

$$+ 3[I^{-2}(X^*) + I^{-2}(X^{*-1})] - 4\,I^{-1}(X^*)$$

whenever $\tau^2 > 0$. □

Remark 2.1. The asymptotic variance τ^2 becomes the limit of $nV_n = \mathrm{Var}\,[n^{1/2} \,.$ $.\,(\mathbb{I}_n^{-1})^S]$ as $n_k \to \infty$, $(k = 1, \ldots, r)$, and it is always lower than the asymptotic variance of $n^{1/2}[I_n^{-1} - I^{-1}(X^*)]$ for the non-stratified random sampling, since

$$-\sum_{k=1}^{r} \frac{N_k}{N} [2 + I^{-1}(k, \cdot) + I^{-1}(\cdot, k)]^2 \leqq -\left[\sum_{k=1}^{r} \frac{N_k}{N} [2 + I^{-1}(\cdot, k) + I^{-1}(k, \cdot)]\right]^2 =$$

$$= -4[I^{-1}(X^*) + 1]^2$$

so that,

$$\tau^2 \leqq -4[I^{-1}(X^*) + 1]^2 + 3[I^{-2}(X^*) + I^{-2}(X^{*-1})] + 2\,I^{-1}(X^*) + 4 =$$

$$= -4[I^{-1}(X^*)]^2 + 3[I^{-2}(X^*) + I^{-2}(X^{*-1})] - 6\,I^{-1}(X^*)$$

and this last expression coincides with that for the asymptotic variance of the unbiased estimator in the non-stratified random sampling ($r = 1$).

This last fact corroborates the advisability of stratifying populations in order to estimate income inequality for large samples in each stratum.

The following result guarantees that in the asymptotic normal distribution of the statistic $n^{1/2}[I_n^{-1} - \boldsymbol{I}^{-1}(\boldsymbol{X}^*)]$, the asymptotic variance τ^2 may be replaced by its analogue estimator τ_n^2 (where τ_n^2 is obtained by replacing the population indices $\boldsymbol{I}^{-1}(\boldsymbol{X}^*)$, $\boldsymbol{I}^{-2}(\boldsymbol{X}^*)$, $\boldsymbol{I}^{-2}(\boldsymbol{X}^{*-1})$, $\boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot)$, $\boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})$, and $\boldsymbol{I}^{-1}(\boldsymbol{k})$ by the corresponding sample indices) without appreciably affecting the accuracy of the approximation. Thus,

Theorem 2.2. The random variable $n^{1/2}[I_n^{-1} - \boldsymbol{I}^{-1}(\boldsymbol{X}^*)]/\tau_n$ is asymptotically distributed according to a standard normal distribution, whenever $\tau_n^2 > 0$ and $\tau^2 > 0$.

Proof. On the basis of the considered sampling, we can state that f_{ik} converges in probability to p_{ik} ($i = 1, \ldots, M$, $k = 1, \ldots, r$) and, consequently, $f_{i.}$ converges in probability to $p_{i.}$. Then, according to the well-known properties of the convergence in probability, we can easily deduce that τ_n^2 converges in probability to τ^2 (and τ_n converges to τ).

As τ is a positive constant value, and $n^{1/2}[I_n^{-1} - \boldsymbol{I}^{-1}(\boldsymbol{X}^*)]$ converges in law to a normal distribution $\mathcal{N}(0, \tau^2)$, it follows that $n^{1/2}[I_n^{-1} - \boldsymbol{I}^{-1}(\boldsymbol{X}^*)]/\tau_n$ converges in law to a standard normal distribution. □

3. THE CHOICE OF SAMPLE SIZE IN ESTIMATING INCOME INEQUALITY FROM COMPLETE DATA

Assume that a margin of error $\varepsilon > 0$ in the estimated income inequality index has been agreed on, and there is a small risk $\alpha > 0$ that we are willing to incur that the actual error is larger than ε when we use the analogue estimator I_n^{-1}.

According to Theorem 2.1, if the asymptotic variance τ^2 were a known positive value, we could use the *asymptotic distribution* of $n^{1/2}[I_n^{-1} - \boldsymbol{I}^{-1}(\boldsymbol{X}^*)]$ to look for the suitable sample size to achieve the preceding requirements. Thus, the value n^* given by

$$n^* = [\lambda_\alpha^2 \tau^2/\varepsilon^2] + 1 \tag{1}$$

(where λ_α is the critical point of the standard normal distribution at the significance level $\frac{1}{2}\alpha$ and [ ] is the greatest integer function), would be adequate for our purposes.

Nevertheless, the population value τ^2 will be unknown in practice and it is in addition impossible to construct a general conservative criterion to choose the sample size similar to that in estimating proportions. In other words, it is not possible to find an upper bound for τ^2 and replace this unknown value by that bound.

In spite of this fact, we could define a non-conservative criterion based on Theorem

2.2, so that the value n^* given by

$$n^* = [\lambda_\alpha^2 \tau_n^2 / \varepsilon^2] + 1 \tag{2}$$

(n = sample size of a previous sample) would be "approximately" adequate for our purposes.

On the other hand, the study of the exact precision of the unbiased estimator $(\mathbb{I}_n^{-1})^S$, we have developed in the First Part of the present paper (Theorems 3.3 and 3.4), allows us now to state another definitively conservative procedure to select the sample size guaranteeing the specified degree of precision determined by the limit error ε and the risk α.

In this way, following *Chebyshev's approach*, we find that such a degree of precision is achieved when we draw at random (and according to a stratified sampling with proportional allocation, independently and with replacement in each stratum) a whole sample of size n^*, where n^* is the integer such that $\mathrm{Var}\,[(\mathbb{I}_n^{-1})^S] \leqq \varepsilon^2\alpha = \varepsilon^*$ for all $n \geqq n^*$. Obviously, the existence of such an integer n^* is confirmed from the fact that $\mathrm{Var}\,[(\mathbb{I}_n^{-1})^S]$ tends to 0 as $n_k \to \infty$ ($k = 1, \ldots, r$), and it may be obtained in practice by solving an inequation of order $s + 2$ (where s = number of strata with different sizes). In particular, when the sizes of all strata coincide, n^* may be obtained by looking for the minimum integer such that

$$A_{\varepsilon^*} n^3 + B_{\varepsilon^*} n^2 + Cn + D > 0 \tag{3}$$

where (if we denote $w_k = N_k/N = n_k/n$),

$$A_{\varepsilon^*} = 2\varepsilon^* \sum_{k=1}^{r} w_k^2$$

$$\begin{aligned} B_{\varepsilon^*} = & -[2\varepsilon^* - 2\sum_{k=1}^{r} w_k^2 \{[\boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})]^2 + [\boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot)]^2 + 2\,\boldsymbol{I}^{-1}(\boldsymbol{X}^*) \,.\, \boldsymbol{I}^{-1}(\boldsymbol{k}) - \\ & - 3[\boldsymbol{I}^{-2}(\boldsymbol{X}^*) + \boldsymbol{I}^{-2}(\boldsymbol{X}^{*-1})] + 4\,\boldsymbol{I}^{-1}(\boldsymbol{X}^*) + 2\,\boldsymbol{I}^{-1}(\boldsymbol{k})\}] \end{aligned}$$

$$\begin{aligned} C = & -[2\sum_{k=1}^{r} \sum_{l=1}^{r} w_k^2 w_l \{[\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l})]^2 + \boldsymbol{I}^{-1}(\boldsymbol{k}) \,.\, \boldsymbol{I}^{-1}(\boldsymbol{l})\} + 2\sum_{k=1}^{r} w_k^2 \{-3[\boldsymbol{I}^{-2}(\cdot, \boldsymbol{k}) + \\ & + \boldsymbol{I}^{-2}(\cdot, \boldsymbol{k}^{-1})] + \boldsymbol{I}^{-1}(\boldsymbol{X}^{*2}) + 2\,\boldsymbol{I}^{-1}(\boldsymbol{X}^*)\} + 2\sum_{k=1}^{r} w_k \{[\boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})]^2 + \\ & + [\boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot)]^2 + 2\,\boldsymbol{I}^{-1}(\boldsymbol{X}^*) \,.\, \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k}) - 3[\boldsymbol{I}^{-2}(\boldsymbol{X}^*) + \boldsymbol{I}^{-2}(\boldsymbol{X}^{*-1})] + \\ & + 4\,\boldsymbol{I}^{-1}(\boldsymbol{X}^*) + 2\,\boldsymbol{I}^{-1}(\boldsymbol{k})\} \end{aligned}$$

$$\begin{aligned} D = & -[-2\sum_{k=1}^{r} \sum_{l=1}^{r} w_k w_l \{[\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l})]^2 + \boldsymbol{I}^{-1}(\boldsymbol{k}) \,.\, \boldsymbol{I}^{-1}(\boldsymbol{l})\} - 2\sum_{k=1}^{r} w_k \{-3[\boldsymbol{I}^{-2}(\cdot, \boldsymbol{k}) + \\ & + \boldsymbol{I}^{-2}(\cdot, \boldsymbol{k}^{-1})] + \boldsymbol{I}^{-1}(\boldsymbol{X}^{*2}) + 2\,\boldsymbol{I}^{-1}(\boldsymbol{X}^*)\} + 2\sum_{k=1}^{r} w_k^2 \{2[\boldsymbol{I}^{-1}(\boldsymbol{k})]^2 - \\ & - 3[\boldsymbol{I}^{-2}(\boldsymbol{k}) + \boldsymbol{I}^{-2}(\boldsymbol{X}^{*-1})] + \boldsymbol{I}^{-1}(\boldsymbol{k}^2) + 2\,\boldsymbol{I}^{-1}(\boldsymbol{k})\}] \end{aligned}$$

In the same way, if we take a random sampling without replacement in each stratum, the required objective may be achieved by looking for the minimum integer n^* such that $\mathrm{Var}\,[(\mathbb{I}_n^{-1})^{SC}] \leqq \varepsilon^2\alpha = \varepsilon^*$ for all $n \geqq n^*$. Obviously, the existence of such an integer n^* is confirmed from the fact that Var $[(\mathbb{I}_n^{-1})^{SC}]$ tends to 0 as $n_k \to N_k$ $(k = 1, \ldots, r)$, and it may be obtained in practice by solving an inequation of order $s + 2$ (where s = number of strata with different sizes). In particular, when the sizes of all strata coincide, n^* may be obtained by looking for the minimum integer such that

(4) $$A_{\varepsilon^*}n^3 + B_{\varepsilon^*}n^2 + Cn + D > 0$$

where

$$\begin{aligned}
A_{\varepsilon^*} = {} & \varepsilon^* \sum_{k=1}^{r} w_k^2 - \sum_{k=1}^{r} w_k^3 \Bigg[\frac{w_k^2}{N_k(N_k - 1)^2 (N_k - 2)(N_k - 3)} \{2N_k(7N_k - 9)\,[\boldsymbol{I}^{-1}(\boldsymbol{k})]^2 - \\
& - 3N_k(5N_k - 7)\,[\boldsymbol{I}^{-2}(\boldsymbol{k}) + \boldsymbol{I}^{-2}(\boldsymbol{k}^{-1})] + (2N_k^2 - 3N_k - 1)\,\boldsymbol{I}^{-1}(\boldsymbol{k}^2) + \\
& + 2(11N_k^2 - 15N_k + 2)\,\boldsymbol{I}^{-1}(\boldsymbol{k})\} + \sum_{l=1}^{r} \frac{w_l^2}{(N_k - 1)(N_l - 1)} \,. \\
& . \{[\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l})]^2 + \boldsymbol{I}^{-1}(\boldsymbol{k}) \,.\, \boldsymbol{I}^{-1}(\boldsymbol{l}) - 3[\boldsymbol{I}^{-2}(\boldsymbol{k}, \boldsymbol{l}) + \boldsymbol{I}^{-2}(\boldsymbol{k}^{-1}, \boldsymbol{l}^{-1})] + \\
& + \boldsymbol{I}^{-1}(\boldsymbol{k}^2, \boldsymbol{l}^2) + 2\,\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l})\} + \frac{1}{N_k - 1} \{[\boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot)]^2 + [\boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})]^2 + \\
& + 2\,\boldsymbol{I}^{-1}(\boldsymbol{X}^*) \,.\, \boldsymbol{I}^{-1}(\boldsymbol{k}) - 3[\boldsymbol{I}^{-2}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-2}(\boldsymbol{k}^{-1}, \cdot)] + 2\,\boldsymbol{I}^{-1}(\boldsymbol{k}) + \\
& + 2\,[\boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})]\} + \frac{2w_k}{(N_k - 1)(N_k - 2)} \{2\,\boldsymbol{I}^{-1}(\boldsymbol{k}) \,.\, [\boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot) + \\
& + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})] - [1 + \boldsymbol{I}^{-1}(\boldsymbol{k})] \,.\, [\boldsymbol{I}^{-1}(\cdot, \boldsymbol{k}^{-1}, \boldsymbol{k}^{-2}) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k}, \boldsymbol{k}^2)] + \\
& + \boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k}) + 2\,\boldsymbol{I}^{-1}(\boldsymbol{k})\} \Bigg]
\end{aligned}$$

$$\begin{aligned}
B_{\varepsilon^*} = {} & -\varepsilon^* + \sum_{k=1}^{r} w_k^2 \Bigg[\frac{w_k^2}{N_k(N_k - 1)^2 (N_k - 2)(N_k - 3)} \{2N_k^2(10N_k^2 - 7N_k - 9)[\boldsymbol{I}^{-1}(\boldsymbol{k})]^2 - \\
& - 3N_k^2(7N_k^2 - 5N_k - 8)\,[\boldsymbol{I}^{-2}(\boldsymbol{k}) + \boldsymbol{I}^{-2}(\boldsymbol{k}^{-1})] + (2N_k^3 + 2N_k^2 - 11N_k + 1) \,. \\
& . \,\boldsymbol{I}^{-1}(\boldsymbol{k}^2) + 2(17N_k^3 - 19N_k^2 - 2N_k - 2)\,\boldsymbol{I}^{-1}(\boldsymbol{k})\} + \sum_{l=1}^{r} \frac{w_l(N_k w_l + N_l w_k + 1)}{(N_k - 1)(N_l - 1)} \,. \\
& . \{[\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l})]^2 + \boldsymbol{I}^{-1}(\boldsymbol{k}) \,.\, \boldsymbol{I}^{-1}(\boldsymbol{l}) + \boldsymbol{I}^{-1}(\boldsymbol{k}^2, \boldsymbol{l}^2) + \\
& +2\,\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l}) - 3[\boldsymbol{I}^{-2}(\boldsymbol{k}, \boldsymbol{l}) + \boldsymbol{I}^{-2}(\boldsymbol{k}^{-1}, \boldsymbol{l}^{-1})]\} + \\
& + \frac{N_k + 1}{N_k - 1} \{2\,\boldsymbol{I}^{-1}(\boldsymbol{X}^*) \,.\, \boldsymbol{I}^{-1}(\boldsymbol{k}) + [\boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot)]^2 + [\boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})]^2 - \\
& - 3[\boldsymbol{I}^{-2}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-2}(\boldsymbol{k}^{-1}, \cdot)] + 2[\boldsymbol{I}^{-1}(\boldsymbol{k}) + \boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})]\} + \\
& + \frac{2w_k(N_k + 1)}{(N_k - 1)(N_k - 2)} \{2\,\boldsymbol{I}^{-1}(\boldsymbol{k}) \,.\, [\boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})] - \\
& - [1 + \boldsymbol{I}^{-1}(\boldsymbol{k})] \,.\, [\boldsymbol{I}^{-1}(\cdot, \boldsymbol{k}^{-1}, \boldsymbol{k}^{-2}) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k}, \boldsymbol{k}^2)] + \\
& + \boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k}) + 2\,\boldsymbol{I}^{-1}(\boldsymbol{k})\} \Bigg]
\end{aligned}$$

$$
\begin{aligned}
C &= \sum_{k=1}^{r} w_k \Biggl[\frac{w_k^2}{(N_k - 1)^2 (N_k - 2)(N_k - 3)} \{-2N_k(N_k^2 + 12N_k - 21)([\boldsymbol{I}^{-1}(\boldsymbol{k})]^2 + \\
&\quad + 12N_k(3N_k - 5)[\boldsymbol{I}^{-2}(\boldsymbol{k}) + \boldsymbol{I}^{-2}(\boldsymbol{k}^{-1})] + (2N_k^3 - 15N_k^2 + 22N_k - 1) . \\
&\quad . \boldsymbol{I}^{-1}(\boldsymbol{k}^2) - 4(2N_k^3 + 3N_k^2 - 5N_k - 1)\, \boldsymbol{I}^{-1}(\boldsymbol{k})\} - \sum_{l=1}^{r} \frac{w_l(N_k N_l w_k + N_k w_l + N_l w_k)}{(N_k - 1)(N_l - 1)} . \\
&\quad . \{[\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l})]^2 + \boldsymbol{I}^{-1}(\boldsymbol{k}) . \boldsymbol{I}^{-1}(\boldsymbol{l}) + \boldsymbol{I}^{-1}(\boldsymbol{k}^2, \boldsymbol{l}^2) + 2\, \boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l}) - 3[\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l}) + \\
&\quad + \boldsymbol{I}^{-1}(\boldsymbol{k}^{-1}, \boldsymbol{l}^{-1})]\} - \frac{N_k}{N_k - 1} \{2\, \boldsymbol{I}^{-1}(X^*) . \boldsymbol{I}^{-1}(\boldsymbol{k}) + [\boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot)]^2 + [\boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})]^2 . \\
&\quad - 3[\boldsymbol{I}^{-2}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-2}(\boldsymbol{k}^{-1}, \cdot)] + 2[\boldsymbol{I}^{-1}(\boldsymbol{k}) + \boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})]\} - \\
&\quad - \frac{2N_k w_k}{(N_k - 1)(N_k - 2)} \{2\, \boldsymbol{I}^{-1}(\boldsymbol{k}) . [\boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})] - [1 + \boldsymbol{I}^{-1}(\boldsymbol{k})] . \\
&\quad . [\boldsymbol{I}^{-1}(\cdot, \boldsymbol{k}^{-1}, \boldsymbol{k}^{-2}) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k}, \boldsymbol{k}^2)] + 2\, \boldsymbol{I}^{-1}(\boldsymbol{k}) + \boldsymbol{I}^{-1}(\boldsymbol{k}, \cdot) + \boldsymbol{I}^{-1}(\cdot, \boldsymbol{k})\} \Biggr] \\
D &= \sum_{k=1}^{r} \frac{N_k w_k}{N_k - 1} \Biggl[\frac{2w_k}{N_k - 1} \{-2N_k[\boldsymbol{I}^{-1}(\boldsymbol{k})]^2 + 3N_k[\boldsymbol{I}^{-2}(\boldsymbol{k}) + \boldsymbol{I}^{-2}(\boldsymbol{k}^{-1})] - \\
&\quad - \boldsymbol{I}^{-1}(\boldsymbol{k}^2) - 2\, \boldsymbol{I}^{-1}(\boldsymbol{k})\} + \sum_{l=1}^{r} \frac{N_l w_l}{N_l - 1} \{[\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l})]^2 + \boldsymbol{I}^{-1}(\boldsymbol{k}) . \boldsymbol{I}^{-1}(\boldsymbol{l}) + \\
&\quad + \boldsymbol{I}^{-1}(\boldsymbol{k}^2, \boldsymbol{l}^2) + 2\, \boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l}) - 3[\boldsymbol{I}^{-1}(\boldsymbol{k}, \boldsymbol{l}) + \boldsymbol{I}^{-1}(\boldsymbol{k}^{-1}, \boldsymbol{l}^{-1})]\} \Biggr]
\end{aligned}
$$

The coefficients in (3) and (4) involve unknown population values, that can be approximated by their respective unbiased estimates in the Appendix of the First Part of this paper. It is worth remarking that the application of Chebyshev's inequality compensates enough the possible error in approximating the population values by means of their unbiased estimates, so that the criterion to choose the sample size remains usually conservative.

Remark 3.1. The integer n^* in Inequations (1), (2), (3) and (4), must be chosen so that $n^* w_k$, $(k = 1, \ldots, r)$ are integer numbers, to allow us to obtain in practice integer sample sizes n_k.

Remark 3.2. Obviously, the procedures to select the sample sizes in (3) and (4) are based on the exact distribution of a random variable, so that they may be applied for small samples and populations (whereas those in (1) and (2) may only be considered for dealing with large samples in each stratum). On the other hand, the criteria based on (1) and (2) always determine smaller sample sizes than those derived from (3) and (4).

4. CONCLUDING REMARKS

A study similar to that in the Second Part of the present paper has been developed [7] for estimating population *entropy* and *diversity*. It would be now interesting to develop another one for estimating the *mutual information* ([6], [8], [9], [10]) in the stratified sampling.

On the other hand, although the research in the First Part of this paper may not easily be accomplished for other income inequality indices, to examine their asymptotic behaviour becomes simple.

In addition, the asymptotic analysis in Section 2 is immediately applicable for defining some useful statistical problems: constructing confidence intervals for the population income inequality, and testing statistical hypotheses concerning this population value, on the basis of the sample income inequality. In virtue of Theorems 2.1 and 2.2, when it is possible to draw large samples in each stratum, these suggested problems may be easily solved.

(Received March 29, 1988.)

REFERENCES

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

[2] C. Caso and M. A. Gil: Estimating income inequality in the stratified sampling from complete data. Part I: The unbiased estimation and applications. Kybernetika *25* (1989), 4, 298—311.

[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] 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.

[6] M. A. Gil and C. Caso: A note on the estimation of the mutual information in the stratified sampling. Metron *45* (1987), 1—2, 295—301.

[7] M. A. Gil and C. Caso: The choice of sample size in estimating entropy according to a stratified sampling. In: Uncertainty and Intelligent Systems. Proc. IPMU Conf. Urbino (Lecture Notes in Computer Science 313), Springer-Verlag, Berlin 1988.

[8] M. A. Gil, M. J. Fernández and I. Martínez: The choice of the sample size in estimating the mutual information. Appl. Math. Comp. (1988) (accepted for publication).

[9] M. A. Gil, R. Pérez and P. Gil: The mutual information. Estimation in the sampling without replacement. Kybernetika *23* (1987), 5, 406—419.

[10] M. A. Gil, R. Pérez and I. Martínez: The mutual information. Estimation in the sampling with replacement. R.A.I.R.O. — Rech. Opér. *20* (1986), 3, 257—268.

[11] Z. Lomnicki and S. Zaremba: The asymptotic distributions of estimators of the amount of transmitted information. Inform. and Control *2* (1959), 260—284.

[12] T. K. Nayak: On diversity measures based on entropy functions. Commun. Statist. - Theor. Meth. *14* (1985), 1, 203—215.

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

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

[15] J. Zvárová: On asymptotic behaviour of a sample estimator of Renyi's information of order α. In: Trans. Sixth Prague Conf. on Inform. Theory, Statist. Dec. Functions Random Processes. Academia, Prague 1973, pp. 919—924.

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