RAIRO. RECHERCHE OPÉRATIONNELLE

M. A. GIL
R. PEREZ
I. MARTINEZ

The mutual information estimation in the sampling with replacement

RAIRO. Recherche opérationnelle, tome 20, nº 3 (1986), p. 257-268
<http://www.numdam.org/item?id=RO_1986__20_3_257_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 20, n° 3, août 1986, p. 257 à 268)

THE MUTUAL INFORMATION ESTIMATION IN THE SAMPLING WITH REPLACEMENT (*)

by M. A. Gil [1], R. Perez [2] and I. Martinez [1]

Abstract. — *The concept of "conditional entropy of order* β *of a random variable with respect to another one" (Z. Daróczy, 1970) leads to the definition of the "amount of information of order* β *conveyed about a random variable by another one", which determines a symmetric measure for both variables. The aim of this paper is to estimate the amount of information of order* $\beta = 2$ *concerning two variables in a population on the basis of a sample drawn from the population. In this way, we first construct an unbiased estimator in the sampling with replacement, and then study its precision. In addition, we discuss the suitability of adopting the measure of order* $\beta = 2$ *instead of the classical amount of information, when we consider large populations.*

Finally, we will corroborate the advantages exhibited by the measure of order $\beta = 2$ *by means of Monte Carlo simulation.*

Keywords: Finite population; sample; Shannon's mutual information; quadratic mutual information; unbiased estimator; sampling with replacement.

Résumé. — *De la notion « entropie conditionnelle de type* β *d'une variable aléatoire sachant la valeur d'une autre variable » (Z. Darózcy, 1970), découle la définition de « quantité d'information de type* β *concernant une variable et contenue dans l'autre » en déterminant une mesure symétrique par rapport aux deux variables. L'objet de cet article est l'estimation de l'information mutuelle de type* $\beta = 2$ *pour deux variables dans une population, à l'aide d'un échantillon prélevé de la population. A ce propos, dans un premier temps nous construisons un estimateur sans biais dans l'échantillonnage avec remplacement, et puis nous examinons sa précision. En outre, nous discutons l'intérêt d'adopter la mesure de type* $\beta = 2$ *au lieu de la mesure classique de Shannon, lorsque nous considérons de grands échantillons.*

Finalement, nous corroborrons les avantages montrés par la mesure de type $\beta = 2$ *avec la méthode, de Monte-Carlo.*

Mots clés : Population finie; échantillon; information mutuelle de Shannon; information mutuelle quadratique; estimateur sans biais; échantillonnage avec remplacement.

1. INTRODUCTION

Consider an experiment involving the observation of two random variables corresponding to measurable characteristics associated with each random choice from a certain finite population. In order to evaluate how much

(*) Received in January 1986.

[1] Departamento de Matemáticas — Facultades de Ciencias, Universidad de Oviedo, 33071 Oviedo (Spain).
[2] Departamento de Estadística y Econometría — Facultad de Económicas, Universidad de Oviedo, 33071 Oviedo (Spain).

information is conveyed about one of the variables by the another one, a usual procedure is to measure it as a reduction in uncertainty.

In this way, let X and Y be two random variables in the considered experiment. When the joint probability distribution of X and Y is known, the uncertainty about the identity of the value of the variable X (or Y), and the uncertainty about X (or Y) when the value of Y is revealed (respectively, when the value of X is revealed) can be quantified by means of a probabilistic uncertainty measure. Then, if $H(X)$ and $H(X/Y)$ (or $H(Y)$ and $H(Y/X)$) denote such uncertainties, "the information conveyed about X by Y" (respectively, the information conveyed about Y by X) can be evaluated by means of the value:

$$\mathscr{I}(X/Y) = H(X) - H(X/Y) \quad \text{(respectively, } \mathscr{I}(Y/X) = H(Y) - H(Y/X)\text{).}$$

In this way, the information conveyed about a random variable by another one is herein intended as the mean decrease in uncertainty about the first variable motivated by the revealment of the value of the second one.

Particularly, if $H(X)$, $H(X, Y)$ and $H(X/Y)$ represent respectively the entropy of order β of X, the joint entropy of order β of (X, Y) (*cf.* Havrda-Charvat [18], Daróczy [11]), and the conditional entropy of order β of X with respect to Y (Daróczy [11]), or the corresponding Shannon's entropies, then:

$$\begin{aligned}\mathscr{I}(X/Y) &= H(X) - H(X/Y) = H(X) + H(Y) - H(X, Y)\\ &= H(Y) - H(Y/X) = \mathscr{I}(Y/X)\end{aligned}$$

that is, the information conveyed about X by Y is the same as the information conveyed about Y by X. For this reason, this symmetric information will be referred as mutual information and will be denoted by $\mathscr{I}(X, Y)$, from now on.

On the other hand, Statistical Inference deals with the drawing of conclusions concerning the variables behavior in a population, on the basis of the variables behavior in a sample from the population. The purpose of the present paper is to estimate the mutual information $\mathscr{I}(X, Y)$ in a finite population given the probabilities of each "value" of (X, Y) in a sample from it when we adopt a random sampling with replacement.

In order to achieve this purpose we will first consider the quadratic mutual information, or mutual information of order $\beta = 2$, and then we will define an unbiased estimator which will be a function of the analogue estimator, the quadratic sample mutual information.

After approaching the precision of the preceding estimation, we will discuss the advantages of the measure of order $\beta = 2$ against Shannon's measure (and the measures of order $\beta \neq 2$) in the sampling with replacement.

These advantages will be finally confirmed by examining a simulated example.

2. PRELIMINARY CONCEPTS

Consider a finite population with N members, and let X and Y be two random variables in the population such that the random vector (X, Y) takes on the pair of real values (x_i, y_j) with joint probabilities p_{ij} $(i = 1, \ldots, M, j = 1, \ldots, M')$, respectively. Let $p_{i.} = \sum_{j=1}^{M'} p_{ij}$ $(i = 1, \ldots, M)$ and $p_{.j} = \sum_{i=1}^{M} p_{ij}$ $(j = 1, \ldots, M')$ the marginal probability distributions of X and Y in the population.

Following Daróczy [11] the mutual information concerning the variables X and Y in the population can be quantified by means of

DEFINITION 2.1: The value $I^2(X, Y)$ defined by:

$$\begin{aligned} I^2(X, Y) &= H^2(X) + H^2(Y) - H^2(X, Y) \\ &= 2\left(1 - \sum_{i=1}^{M} p_{i.}^2\right) + 2\left(1 - \sum_{j=1}^{M'} p_{.j}^2\right) - 2\left(1 - \sum_{i=1}^{M} \sum_{j=1}^{M'} p_{ij}^2\right) \\ &= 2\left(1 - \sum_{i=1}^{M} p_{i.}^2 - \sum_{j=1}^{M'} p_{.j}^2 + \sum_{i=1}^{M} \sum_{j=1}^{M'} p_{ij}^2\right) \end{aligned}$$

is called *quadratic population mutual information* concerning X and Y.

The analogue estimator of this value is introduced in an immediate way. If we consider a sample of n members drawn at random, denoted by (x, y), taking on the pairs of values (x_i, y_j) with relative frequencies $f_{ij}(x, y)$ $(i = 1, \ldots, M, j = 1, \ldots, M')$, respectively, the mutual information concerning X and Y in the sample can be quantified by means of

DEFINITION 2.2: The value $I_n^2(x, y)$ defined by:

$$\begin{aligned} I_n^2(x, y) &= H_n^2(x) + H_n^2(y) - H_n^2(x, y) \\ &= 2\left(1 - \sum_{i=1}^{M} [f_{i.}(x)]^2\right) + 2\left(1 - \sum_{j=1}^{M'} [f_{.j}(y)]^2\right) \\ &\quad - 2\left(1 - \sum_{i=1}^{M} \sum_{j=1}^{M'} [f_{ij}(x, y)]^2\right) \\ &= 2\left(1 - \sum_{i=1}^{M} [f_{i.}(x)]^2 - \sum_{j=1}^{M'} [f_{.j}(y)]^2 + \sum_{i=1}^{M} \sum_{j=1}^{M'} [f_{ij}(x, y)]^2\right) \end{aligned}$$

$$\left(\text{being } f_{i.}(x) = \sum_{j=1}^{M'} f_{ij}(x, y),\ f_{.j}(y) = \sum_{i=1}^{M} f_{ij}(x, y),\quad i = 1, \ldots, M,\ \ j = 1, \ldots, M'\right)$$

is called *quadratic sample mutual information* concerning X and Y.

REMARK 2.1: It is worth emphasizing that the quadratic entropy (H^2) in the preceding definitions has the qualitative signification and essential properties in Shannon's entropy, for quantifying probabilistic uncertainty. In the same way, the quadratic mutual information has the qualitative signification and essential properties in Shannon's mutual information (recalled in Section 6) for quantifying the information conveyed about one variable or experiment by another one.

REMARK 2.2: The quadratic mutual information above defined has some interesting applications which allows us to interpretate this concept in different fields.

As a first application of this information, we consider the problem of measuring the ecological diversity of a finite population under a classification process X dividing it into M classes or species, when the population is subjected to an additional separate classification process Y that divides it into M' classes. If the diversity under a classification within each class determined by the other one is measured by means of the Gini-Simpson index ([5], [6], [7], [25], [26]), then $I^2(X, Y)$ quantifies the mean decrease in diversity (or, the mean increase in concentration) under the X-classification motivated by the adoption of the additional classification process Y, and conversely, the mean decrease in diversity under the Y-classification motivated by the X-classification.

A second application of the quadratic mutual information is the evaluation of the quadratic information processed by a discrete constant channel (and, the capacity of type $\beta = 2$ of this channel) with input alphabet X, characterized by M symbols, and output alphabet Y, characterized by M' symbols (*cf.* [11]). A similar application is given by the evaluation of the informations processed and transmitted by a questionnaire having M questions and M' answers ([3],[13], [24]).

3. AN UNBIASED ESTIMATOR OF THE POPULATION MUTUAL INFORMATION IN THE SAMPLING WITH REPLACEMENT

When we consider the notations in Section 2, the random vectors $(nf_{11}, \ldots, nf_{MM'})$, $(nf_{1.}, \ldots, nf_{M.})$ and $(nf_{.1}, \ldots, nf_{.M'})$ have multinomial distributions with parameters $(n, p_{11}, \ldots, p_{MM'})$, $(n, p_{1.}, \ldots, p_{M.})$ and $(n, p_{.1}, \ldots, p_{.M'})$, respectively. Consequently, the expected value of the analogue estimator in Definition 2.2, over all samples (x, y) of size n in a random sampling with replacement, is given by:

$$E(I_n^2) = 2\left[1 - \sum_{i=1}^{M} E(f_{i.}^2) - \sum_{j=1}^{M'} E(f_{.j}^2) + \sum_{i=1}^{M} \sum_{j=1}^{M'} E(f_{ij}^2)\right] = \frac{n-1}{n} I^2(X, Y)$$

Therefore, the analogue estimator I_n^2 is asymptotically unbiased. In addition, it allows us to construct an unbiased estimator of $I^2(X, Y)$, since whatever the size n may be we have.

THEOREM 3.1: *Let (X, Y) be a random vector in a finite population taking on the pairs of values (x_i, y_j) $(i = 1, \ldots, M, j = 1, \ldots, M')$. In the random sampling with replacement from this population, the estimator $\hat{I}_n^2$ allocating to each sample (x, y) of n members the value $\hat{I}_n^2(x, y) = n\, I_n^2(x, y)/(n-1)$ is an unbiased estimator of the quadratic population mutual information concerning X and Y.* ■

Theorem 3.1 suggests the introduction of the following concept:

DEFINITION 3.1: The estimator $\hat{I}_n^2$ allocating to each sample (x, y) of size n with replacement the value $\hat{I}_n^2(x, y) = n\ I_n^2(x, y)/(n-1)$ is called *quadratic sample mutual quasi-information* concerning X and Y.

4. EXACT PRECISION OF THE UNBIASED ESTIMATOR

In order to evaluate the exact precision of the quadratic sample mutual quasi-information in estimating the quadratic population mutual information, we now measure the mean square error of that estimator. As it is an unbiased estimator, the corresponding mean square error is given by the estimator variance.

THEOREM 4.1: *Let (X, Y) be a random vector in a finite population taking on the pairs of values (x_i, y_j) $(i = 1, \ldots, M, j = 1, \ldots, M')$. If $\hat{I}_n^2$ is the quadratic mutual quasi-information for a random sample of size n with replacement, then its variance:*

$$V(\hat{I}_n^2) = \Big\{(6-4n)[I^2(X, Y)]^2 - 12(n-2)I^3(X, Y) + 4(4n-7)I^2(X, Y) + 32(n-2)\sum_{i=1}^{M}\sum_{j=1}^{M'} p_{ij}(p_{i.} - p_{ij})(p_{.j} - p_{ij})\Big\}/n(n-1)$$

$$\left(\textit{being } I^3(X, Y) = H^3(X) + H^3(Y) - H^3(X, Y) = \frac{4}{3}\left(1 - \sum_{i=1}^{M} p_{i.}^3 - \sum_{j=1}^{M'} p_{.j}^3 + \sum_{i=1}^{M}\sum_{j=1}^{M'} p_{ij}^3\right) \textit{the mutual information of order } \beta = 3 \textit{ concerning } X \textit{ and } Y\right).$$

Proof: As the random vectors $(nf_{11}, \ldots, nf_{MM'})$, $(nf_{1.}, \ldots, nf_{M.})$ and $(nf_{.1}, \ldots, nf_{.M'})$ have multinomial distributions with parameters $(n, p_{11}, \ldots, p_{MM'})$, $(n, p_{1.}, \ldots, p_{M.})$ and $(n, p_{.1}, \ldots, p_{.M'})$, respectively, we have:

$$V(\hat{I}_n^2) = \Big\{4(6-4n)\Big(\sum_{i=1}^{M} p_{i.}^2 + \sum_{j=1}^{M'} p_{.j}^2 - \sum_{i=1}^{M}\sum_{j=1}^{M'} p_{ij}^2\Big)^2$$

$$+ 16(n-2)\left(\sum_{i=1}^{M} p_{i.}^3 + \sum_{j=1}^{M'} p_{.j}^3 - \sum_{i=1}^{M}\sum_{j=1}^{M'} p_{ij}^3\right)$$

$$+ 8\left(\sum_{i=1}^{M} p_{i.}^2 + \sum_{j=1}^{M'} p_{.j}^2 - \sum_{i=1}^{M}\sum_{j=1}^{M'} p_{ij}^2\right)$$

$$+ 32(n-2)\sum_{i=1}^{M}\sum_{j=1}^{M'} p_{ij}(p_{i.}-p_{ij})(p_{.j}-p_{ij})\Bigg\}/n(n-1)$$

which accounts for proving the theorem. ■

REMARK 4.1: The expression $\sum_{i=1}^{M}\sum_{j=1}^{M'} p_{ij}(p_{i.}-p_{ij})(p_{.j}-p_{ij})$ in $V(\hat{I}_n^2)$ (Theorem 4.1) cannot be stated only in terms of mutual informations of order β, but it may be further stated in terms of other measures concerning another basic concept in the Information Theory and Statistics: the inaccuracy. (In addition, the entropy of order β is the inaccuracy of order β of a probability distribution with respect to itself). So, $\sum_{i=1}^{M}\sum_{j=1}^{M'} p_{ij}(p_{i.}-p_{ij})(p_{.j}-p_{ij}) = \frac{I^2(X, Y)}{2}[1 - H_2(P; Q)/2]$, being $H_2(P; Q)$ the inaccuracy of order $\beta = 2$ of P with respect to Q, where P and Q denote the probability distributions $\{p_{ij}\}$ and $\left\{\frac{(p_{i.}-p_{ij})(p_{.j}-p_{ij})}{I^2(X, Y)/2}\right\}$, respectively, [26].

REMARK 4.2: Theorem 4.1 implies that zero is the limit of $V(\hat{I}_n^2)$ as $n \to \infty$. Therefore, zero is the limit of the standard error $[V(\hat{I}_n^2)]^{1/2}$ as $n \to \infty$.

Remark 4.2 allows us to conclude that for a large sample the standard error of the estimator $\hat{I}_n^2$ is small. In addition, we can verify that the greater the size of the sample, the lower its mean square error is, as we now prove in the following:

THEOREM 4.2: *Whatever the sample size n may be, then:*

$$V(\hat{I}_n^2) - V(\hat{I}_{n-1}^2) = -[(n-2)(n-3)V(\hat{I}_{n-2}^2) + 2V(\hat{I}_2^2)]/n(n-1)(n-2). \quad ■$$

5. ESTIMATED PRECISION OF THE UNBIASED ESTIMATOR

As the mean square error of $\hat{I}_n^2$ involves population probabilities of the variable values, this error will not be known in practice. However, this error can be estimated from the considered sample. The following result states an unbiased estimator of $V(\hat{I}_n^2)$ defined on the basis of the analogue estimates of $I^2(X, Y)$, $I^3(X, Y)$ and $\sum_{i=1}^{M}\sum_{j=1}^{M'} p_{ij}(p_{i.}-p_{ij})(p_{.j}-p_{ij})$.

THEOREM 5.1: *Let* (X, Y) *be a random vector in a finite population taking on the pairs of values* (x_i, y_j) $(i = 1, \ldots, M, j = 1, \ldots, M')$ If $v(\hat{I}_n^2)$ *is the estimator allocating to each sample* (x, y) *of* n *members in a random sampling with replacement the value given by:*

$$v(\hat{I}_n^2)(x, y) = \Big\{ n^2(6 - 4n)[I_n^2(x, y)]^2 - 12n^2(n-1)I_n^3(x, y) + 4n(n-1)(4n-1)I_n^2(x, y) + 32n^2(n-1)\sum_{i=1}^{M}\sum_{j=1}^{M'} f_{ij}(x, y)(f_{i.}(x) - f_{ij}(x, y))(f_{.j}(y) - f_{ij}(x, y))\Big\}/(n-1)^2(n-2)(n-3)$$

$\Big($being $I_n^3(x, y) = H_n^3(x) + H_n^3(y) - H_n^3(x, y) = \dfrac{4}{3}\Big\{1 - \sum_{i=1}^{M}\{f_{i.}(x)]^3 - \sum_{j=1}^{M'}[f_{.j}(y)]^3 + \sum_{i=1}^{M}\sum_{j=1}^{M'}[f_{ij}(x, y)]^3\Big\}$

the mutual information of order $\beta = 3$ concerning X and Y in the sample $(x, y)\Big)$,

then, $v(\hat{I}_n^2)$ is an unbiased estimator of $V(\hat{I}_n^2)$. ■

REMARK 5.1: Theorem 5.1 connects n with an unbiased estimate of the precision of $\hat{I}_n^2$. With such a connection one could readily estimate the suitable size for estimating the quadratic population mutual information (and consequently, the mean decrease in diversity, the information processed by a channel, and so on), by means of the quadratic sample mutual quasi-information, with a desired degree of precision. The estimation of this suitable size could be accomplished either by using a previous sampling from the population (in order to approximate $I_n^2(x, y)$, $I_n^3(x, y)$, $f_{ij}(x, y)$, $f_{i.}(x)$ and $f_{.j}(y)$), or by using a sequential sampling.

REMARK 5.2: The size and finiteness of the population are irrelevant in the main results of the present paper (e.g., Theorems 3.1, 4.1, 4.2 and 5.1).

6. ADVANTAGES OF THE QUADRATIC MUTUAL INFORMATION AGAINST THE SHANNON'S MUTUAL INFORMATION

Consider the finite population, the associated variables X and Y, a generic sample of size n with replacement, (x, y), and the notations in the preceding sections.

Following Shannon, the mutual information concerning the variables X and Y in the population is quantified by means of:

DEFINITION 6.1: The value $I(X, Y)$ defined by:

$$\begin{aligned} I(X, Y) &= H(X) + H(Y) - H(X, Y) \\ &= -\sum_{i=1}^{M} p_{i.} \log_2 p_{i.} - \sum_{j=1}^{M'} p_{.j} \log_2 p_{.j} + \sum_{i=1}^{M} \sum_{j=1}^{M'} p_{ij} \log_2 p_{ij} \\ &= \sum_{i=1}^{M} \sum_{j=1}^{M'} p_{ij} \log_2 \frac{p_{ij}}{p_{i.} p_{.j}} \end{aligned}$$

is called *Shannon's population mutual information* concerning X and Y.

The analogue estimator of this value is defined for the random sample (x, y) as follows:

DEFINITION 6.2: The value $I_n(x, y)$ defined by:

$$\begin{aligned} I_n(x, y) &= H_n(x) + H_n(y) - H_n(x, y) \\ &= -\sum_{i=1}^{M} f_{i.}(x) \log_2 f_{i.}(x) - \sum_{j=1}^{M'} f_{.j}(y) \log_2 f_{.j}(y) \\ &+ \sum_{i=1}^{M} \sum_{j=1}^{M'} f_{ij}(x, y) \log_2 f_{ij}(x, y) \\ &= \sum_{i=1}^{M} \sum_{j=1}^{M'} f_{ij}(x, y) \log_2 \frac{f_{ij}(x, y)}{f_{i.}(x) f_{.j}(y)} \end{aligned}$$

is called *Shannon's sample mutual information* concerning X and Y.

When we examine the expected value of I_n over all samples of size n in a random sampling with replacement, we obtain:

$$E(I_n) = \sum_{i=1}^{M} \sum_{j=1}^{M'} E(f_{ij} \log_2 f_{ij}) - \sum_{i=1}^{M} E(f_{i.} \log_2 f_{i.}) - \sum_{j=1}^{M'} E(f_{.j} \log_2 f_{.j}).$$

Nevertheless, an exact relation, irrespective of the variables X and Y, cannot be established either between $E(f_{ij} \log_2 f_{ij})$ and $E(f_{ij}) \log_2 E(f_{ij}) = p_{ij} \log p_{ij}$, $E(f_{i.} \log_2 f_{i.})$ and $E(f_{i.}) \log_2 E(f_{i.}) = p_{i.} \log_2 p_{i.}$, or $E(f_{.j} \log_2 f_{.j})$ and $E(f_{.j}) \log_2 E(f_{.j}) = p_{.j} \log_2 p_{.j}$.

Consequently, an unbiased estimator of I irrespective of the variables X and Y, cannot be immediately defined from I_n.

In the same way, if I^β and I_n^β denote respectively the population and sample mutual information of order β ($\beta \neq 1, 2$) (Daróczy, [11]), to state exact relations between $E(I_n^\beta)$ and I^β is either impossible or very complex.

7. SIMULATED EXAMPLE

In order to corroborate the results in the preceding sections, we now analyze an example when we assume the population distribution is known (although such results will be useful in practice when that distribution is unknown). Consider the bivariate normal distribution $N(\{\mu_r\}, \|\sigma_{rs}\|)$, $r = 1, 2$, $s = 1, 2$, with $\mu_1 = \mu_2 = 0$, $\sigma_{11} = \sigma_{22} = 1$, $\sigma_{12} = \sigma_{21} = 2^{-1/2}$, where the values have been grouped according to 40 intervals, $[-4, -3.8]$, $[-3.8, -3.6]$, $\ldots$, $[3.6, 3.8]$ and $[3.8, 4]$ for each of the marginal distributions.

Let (X, Y) be the random vector whose values are the 40×40 pairs determined by the mid-points in both groupings, and so that the probability of each of these pairs is given by that of the corresponding product interval for the bivariate normal distribution.

Then, the Shannon's population mutual information is given by:

$$I(X, Y) = 0.494{,}625{,}164{,}5$$

and the quadratic population mutual information is given by:

$$I^2(X, Y) = 1.783{,}644{,}870{,}0.$$

For the sample size $n = 5000$, 100 samples of size n have been generated from the joint distribution of X and Y by Monte Carlo simulation. For each sample, I_n and I_n^2 have been calculated.

Being n a very great size, we can approach $\hat{I}_n^2$ by means of I_n^2 and compare the estimations of $I(X, Y)$ and $I^2(X, Y)$ given, respectively, by their analogue estimates. This comparison is now stated by evaluating the mean of the estimates of $I(X, Y)$, which has been given by:

$$\bar{I}_n = 0.563{,}272{,}295{,}8$$

and the mean of the estimates of $I^2(X, Y)$, which has been given by:

$$\bar{I}_n^2 = 1.784{,}541{,}862{,}4.$$

In addition, the mean square error of the 100 estimates of $I(X, Y)$ has been given by:

$$\overline{(I_n - I(X, Y))^2} = 0.004{,}720{,}211{,}4$$

and the mean square error of the 100 estimates of $I^2(X, Y)$ by:

$$\overline{(I_n^2 - I^2(X, Y))^2} = 0.000{,}000{,}941{,}2.$$

On the other hand, the ratio (biais/standard deviation) for the 100 estimates of $I(X, Y)$ equals to 24.703,763,337,4, and for the 100 estimates of $I^2(X, Y)$ equals to 2.426,964,261,3.

Consequently, the estimation of $I(X, Y)$ by means of I_n is "less suitable" than the estimation of $I^2(X, Y)$ by means of I_n^2.

8. CONCLUDING REMARKS

The study in this paper could be developed for the random sampling without replacement and the stratified random sampling, which would provide greater precisions.

In the same way, the estimation of the population mutual information may be examined for the case when the adopted sampling is not random.

On the other hand, a similar study for estimating the uncertainty associated with a variable in a finite population has been accomplished in [23]. It should be emphasized the existence of many analogies between the results in [23] and the present paper.

Another interesting similar study would be determined by the estimation of the unquietness associated with a random variable in a finite population, which could be applied to the estimation of the income inequality in a large population, [22], and [8], [10], [12], [14], [15], [16], [28], [30], [31].

Finally, the results we have just expounded migth be used for estimating the information conveyed by a sample about its corresponding population (with respect to a random variable).

ACKNOWLEDGEMENT

The authors are grateful to Professor Robert Fortet for his suggestions in connection with this paper.

APPENDIX 1

Moments of the multinomial distribution

Let $(n_1, \ldots, n_M)$ be a random vector with multinomial distribution, where n, $p_1, \ldots, p_M \left(\sum_{i=1}^{M} p_i = 1\right)$ are the parameters. Then,

$$E(n_i) = np_i$$
$$E(n_i^2) = np_i[(n-1)p_i + 1]$$

$$E(n_i^3) = np_i[(n-1)(n-2)p_i^2 + 3(n-1)p_i + 1]$$
$$E(n_i^4) = np_i[(n-1)(n-2)(n-3)p_i^3 + 6(n-1)(n-2)p_i^2 + 7(n-1)p_i + 1]$$
$$E(n_i n_j) = n(n-1)p_i p_j, \qquad i \neq j$$
$$E(n_i^2 n_j^2) = n(n-1)p_i p_j[(n-2)(n-3)p_i p_j + (n-2)(p_i + p_j) + 1], \qquad i \neq j$$
$$E(n_i^3 n_j) = n(n-1)p_i p_j[(n-2)(n-3)p_i^2 + 3(n-2)p_i + 1], \qquad i \neq j$$
$$E(n_i^2 n_j n_k) = n(n-1)(n-2)p_i p_j p_k[(n-3)p_i + 1], \qquad j \neq i, \qquad k \neq i, j$$
$$E(n_i n_j n_k n_l) = n(n-1)(n-2)(n-3)p_i p_j p_k p_l, \qquad j \neq i, \qquad k \neq i, j, \qquad l \neq i, j, k.$$

REFERENCES

1. J. Aczel and Z. Daroczy, *On Measures of Information and Their Characterizations,* Academic Press (New York), 1975.
2. J. Aczel and Pl. Kannapan, *A Mixed Theory of Information.* III. Inset Entropies of Degree β, Inf. and Contr., 39, 1978, 315-322.
3. N. L. Aggarwal, Y. Cesari and C. F. Picard, *Propriétés de branchement liées aux questionnaires de Campbell et à l'information de Renyi,* C.R. Acad. Sc. Paris, Série A, t.275, 437-440, 1972.
4. F. Azorin, *Curso de Muestreo y Applicaciones,* Ed. Aguilar (Madrid), 1972.
5. F. Azorin, *Diversity and Uncertainty,* Proc. F.I.S.A.L.-83 (Palma de Mallorca), 1983, 37-44.
6. T. N. Bhargava and V. R. R. Uppuluri, *On an axiomatic derivation of Gini diversity with applications,* Metron, Vol. XXXIII, N.1-2, 1985.
7. T. N. Bhargava and V. R. R. Uppuluri, *Sampling Distribution of Gini's index of diversity,* Appl. Math. Comput., 3, 1-24, 1977.
8. F. Bourguignon, *Decomposable income inequality measures,* Econometrica, 47, 901-920, 1979.
9. W. G. Cochran, *Sampling Techniques,* John Wiley (New York), 1967.
10. F. A. Cowell, *On the structure of additive inequality measures,* Review of Economic Studies, 47, 521-531, 1980.
11. Z. Daroczy, *Generalized Information Functions,* Inf. and Contr., 16, 36-51, 1970.
12. W. Eichhorn and W. Gehrig, *Measurement of Inequality in Economics,* Modern Applied Mathematics Optimization and Operations Research, Ed. B. Korte, North Holland (Amsterdam), 1982.
13. H. Emptoz, *Informations utiles et pseudoquestionnaires,* These, Université Claude-Bernard (Lyon), 1976.
14. M. A. Gil, *Estudio de una medida para la incertidumbre correspondiente a las utilidades,* Trab. de Est. e Inv. Oper., Vol. 32, 3, 45-66, 1981.
15. M. A. Gil, *Criterion of Maximizing the Expected Quietness (invariant by homotheties in relation to the utilities),* R.A.I.R.O.-Rech. Opér., 16, 319-331, 1982.
16. M. A. Gil, *Mixed Criterion of Expected Utility and Quietness (invariant by homotheties with respect to the utilities),* Statistica, Anno XLII, n.1, 21-37, 1982.
17. P. Gil *Teoria Matemática de la Información,* Ed. ICE (Madrid), 1981.

18. J. Havrda and F. Charvat, *Quantification method of classification processes,* Kybernetica, 3, 30-35, 1967.
19. Sir M. Kendall and A. Stuart, *The Advanced Theory of Statistics,* Charles Griffin Co. Ltd. (London), 1977-83.
20. A. M. Mathai and P. N. Rathie, *Basic Concepts in Information Theory and Statistics,* Wiley Eastern Limited (New Delhi), 1975.
21. O. Onicescu, *Energie Informationnelle,* C. R. Acad. Sci. Paris, Série A, t. 263, 1966, 841-842.
22. R. Perez, *Estimación de la incertidumbre, la incertidumbre ùtil y la inquietud en poblaciones finitas. Una aplicación a las medidas de desigualdad.* Sesión Científica de la Real Acad. Ciencias de Madrid, 10 Abril, 1985.
23. R. Perez, M. A. Gil and P. Gil, *Estimating the Uncertainly associated with a Variable in a Finite Population,* Kybernetes, 15 (3), 1986.
24. C. F. Picard, *Graphes et Questionnaires,* Tome 2, Gauthier-Villars Ed. (Paris), 1972.
25. E. C. Pielou, *An Introduction to Mathematical Diversity,* Wiley-Interscience (New York), 1969.
26. P. N. Rathie and Pl. Kannappan, *An inaccuracy function of the type* β, Ann. Inst. Statist. Math., 25, 205-214, 1973.
27. C. E. Shannon, *A Mathematical Theory of Communication,* Bell. System. Tech., J. 27, 379-423, 1948.
28. A. F. Shorrocks, *The class of additively decomposable inequality measures,* Econometrica, 48, 613-625, 1980.
29. E. H. Simpson, *Measurement of Diversity,* Nature, 163, 688, 1949.
30. H. Theil, *Economics and Information Theory,* North Holland (Amsterdam), 1967.
31. D. Zagier, *On the decomposability of the Gini coefficient and other indices of inequality.* Discussion Paper No. 108, Projectgruppe Theoretische Modelle, Universität Bonn, 1983.