KYBERNETIKA — VOLUME 27 (1991), NUMBER 6

COMPARING THE SUITABILITY OF TWO FACTORS FOR STRATIFICATION IN ESTIMATING DIVERSITY

MARÍA ANGELES GIL

When one tries estimate the "diversity of species" in an ecological community on the basis of a large sample, one can guarantee a gain in precision from stratified random over simple random sampling. In addition, many ecological communities to which that estimation is applied may be naturally stratified by different factors. Should this be the case, one could compare two different "factors for stratification" to conclude if they exert the same influence on the distribution of species. This comparison will be formalized in this paper by developing a procedure to test the hypothesis that the "mutual information" between the classification process determining different species and one of the factors coincides with that between the classification process and the other factor. Then, procedures to test the corresponding one-sided hypotheses will be discussed. Finally, the application of those procedures will be illustrated by means of a practical example.

1. INTRODUCTION

Species diversity in community ecology is usually thought of (cf. Ludwig & Reynolds [18]) as being composed of two components: the number of species in the community and the eveness (that refers to how the species abundances are distributed among the species). There are many diversity indices suggested in the literature, although those based on well-known entropy measures (such as Shannon's entropy [26]), and Havrda-Charvát's non-additive entropies of order α, [15]) are the most commonly employed (see, for instance, Good [14], Margalef [19], Basharin [1], Bhargava & Doyle [2], Pielou [22], Bhargava & Uppuluri [3], Routledge [25], Holtage [16], Patil & Taille [21], Rao [23], Nayak [20]).

When a community is too large to be censused, the diversity of species in it with respect to a classification process must be estimated from a sample. In this situation, we have verified (see, Gil [10]) that a gain in precision from stratified random sampling (with proportional allocation in each stratum and independently in different strata) over simple random sampling may be guaranteed, if large samples from the community are available. On the basis of this argument, and to obtain estimates as precise as possible, it is useful to take into account that many real-life communities to which that estimation is applied arise naturally stratified by means of different factors. (Thus, for instance, in the measurement of the diversity of species of trees in a forest, the trees could be stratified according to the range of their heights above ground or by a geological classification regarding the type of terrain).

When there is more than one natural factor to stratify a community, it would be interesting to find the one exerting the strongest influence of the distribution of species. In this sense, we are now going to analyze the problem of testing whether two different factors for stratification result in the same influence on the distribution of species in the community, on the basis of a large sample drawn at random from it.

To approach this problem, we first take into account that in Sampling Theory concerning the estimation of the population mean or total of a random variable, an "ideal factor for stratification" would divide the population so that each stratum would contain individuals taking on the same variable values (or very close values). This idea is often formalized by saying that an ideal factor for stratification would present a high value for the *Pearson correlation coefficient* (or other association coefficient) between the random variable and itself. Nevertheless, the diversity depends only on the abundances of the possible categories for the classification process, but in no way depends on what these categories are, and hence an "ideal factor for stratification" in estimating diversity would divide the population so that each stratum would contain a dominant category and a small number of different categories. This last criterion may be formalized by saying that an ideal factor for stratification would present a high value for the *mutual information* (or average decrease in the diversity associated with the classification process, due to the knowledge of the stratum) between the classification process and itself (see, Remark 2.2 below).

Remark 1.1. An alternative way to measure the suitability of a factor for stratification would be given by the "gain in precision from stratified random sampling over simple random sampling". The inconveniences in considering this measure arise because of the fact that the gain in precision depends on the allocation according to which the stratified sampling is carried out. On the other hand, the expression for that gain in precision becomes, for all the allocations, more complex to handle than that for the mutual information index.

The preceding assertions motivate us to develop in this paper a procedure to test the null hypothesis that "the mutual information between the classification process and one of the factors equals the mutual information between the classification process and the other factor". Then, procedures for testing the corresponding one-sided hypotheses will be defined and, finally, we will examine a practical example using those techniques.

The study in this paper will refer to a sampling study based on one sample. Studies based on two independent samples (when available) would be immediately derived from asymptotic analyses in [17], [29] and [30]. In other words, the main contribution of this paper is due to the connection established between diversity and mutual information, the use of this connection to stratify populations in estimating diversity, and the possibility of comparing two factors for stratification on the basis of only one sample.

We next establish the basic concepts and results for this study.

2. PRELIMINARY CONCEPTS

Consider a community with N individuals that can be classified according to a classification process X into M species, $x_1, \ldots, x_M$. Let Y be a factor for stratification dividing the community into r non-overlapping strata, $y_1, \ldots, y_r$ and let Z be a factor for stratification dividing the community into s non-overlapping strata, $z_1, \ldots, z_s$. If p_{ijk} denotes the relative abundance of the species x_i belonging to both, the stratum y_j and the stratum z_k $(i = 1, \ldots, M,\ j = 1, \ldots, r,\ k = 1, \ldots, s)$ the diversity of species in the community may be quantified by

Definition 2.1. The value $H^\alpha(X)$ defined by

$$H^\alpha(X) = (\alpha - 1)^{-1}\left[1 - \sum_{i=1}^{M}\left(\sum_{j=1}^{r}\sum_{k=1}^{s} p_{ijk}\right)^\alpha\right]$$

if $\alpha > 0$, $\alpha \neq 1$, is called *Havrda-Charvát's population diversity index of order* α associated with X, and the value

$$H^1(X) = -\sum_{i=1}^{M}\left(\sum_{j=1}^{r}\sum_{k=1}^{s} p_{ijk}\right)\log\left(\sum_{j=1}^{r}\sum_{k=1}^{s} p_{ijk}\right)$$

if $\alpha = 1$, is called *Shannon's population diversity index* associated with X.

On the other hand, the mutual information between the classification process X and each of the factors Y and Z in the community may be evaluated (cf., Daget & Godron [7]) by

Definition 2.2. The value $I^\alpha(X; Y)$ defined by

$$I^\alpha(X; Y) = (\alpha - 1)^{-1}\left[1 - \sum_{i=1}^{M}\left(\sum_{j=1}^{r}\sum_{k=1}^{s} p_{ijk}\right)^\alpha - \sum_{j=1}^{r}\left(\sum_{i=1}^{M}\sum_{k=1}^{s} p_{ijk}\right)^\alpha + \right.$$
$$\left. + \sum_{i=1}^{M}\sum_{j=1}^{r}\left(\sum_{k=1}^{s} p_{ijk}\right)^\alpha\right]$$

if $\alpha > 0$, $\alpha \neq 1$, is called *population mutual information index of order* α concerning X and Y, and

$$I^1(X; Y) = \sum_{i=1}^{M}\sum_{j=1}^{r}\sum_{k=1}^{s} p_{ijk}\log\frac{\left(\sum_{m=1}^{s} p_{ijm}\right)}{\left(\sum_{h=1}^{M}\sum_{m=1}^{s} p_{hjm}\right)\left(\sum_{l=1}^{r}\sum_{m=1}^{s} p_{ilm}\right)}$$

if $\alpha = 1$, is called *Shannon's population mutual information index* concerning X and Y.

The values $I^{\alpha}(X; Z)$ and $I^{1}(X; Z)$ would be defined in a similar way.

Remark 2.1. The preceding measures of mutual information have been constructed on the basis of the Havrda-Charvát non-additive entropies of order α, for $\alpha > 0$, $\alpha \neq 1$, and Shannon's entropy, for $\alpha = 1$, and the corresponding conditional entropies (Daróczy [8]). Properties of these measures, guaranteeing their suitability to quantify the mutual information between two random variables have been examined in previous papers (Emptoz [9], Gil et al. [12]). The Havrda-Charvát measures are based on the well-known Rényi measures [24], but although a comparison trough the first ones is equivalent to a comparison trough the last ones, Havrda-Charvát's measures are more operative for the problem of estimation we are considering.

Remark 2.2. Indices of mutual information in Definition 2.2 are defined so that $I^{\alpha}(X; Y) = H^{\alpha}(X) - H^{\alpha}(X \mid Y)$, where $H^{\alpha}(X \mid Y)$ is the conditional entropy of order α defined by $\sum_j p(y_j)^{\alpha} H^{\alpha}(X \mid y_j)$ (where $p(y_j) = \sum_i \sum_k p_{ijk}$, and $H^{\alpha}(X \mid y_j)$ is the entropy of order α of the conditional distribution of X given $Y = y_j$). Then, $I^{\alpha}(X; Y) \leqq H^{\alpha}(X)$, whatever Y may be, and hence a high value for the mutual information means a low value for each of the $H^{\alpha}(X \mid y_j)$'s. Furthermore, if Y is a factor for stratification, a low value of $H^{\alpha}(X \mid y_j)$ indicates that in stratum y_j there is a dominant category of X and a small number of different categories, and this fact justifies the criterion of stratification based on mutual information indices.

When the community is too large to be censused, the diversity associated with X, and the mutual information between X and each of the factors for stratification, must be estimated on the basis of a sample drawn from the community. Let us assume that a sample of n individuals is drawn at random and with replacement from the whole community. Let f_{ijk} denote the relative abundance of the species x_i belonging to both, the stratum y_j and the stratum z_k ($i = 1, \ldots, M$, $j = 1, \ldots, r$, $k = 1, \ldots, s$), for the sample. Then, the diversity of species in the sample could be quantified by means of the analogue estimator of the population index, $\eta^{\alpha}(X)$, (obtained by replacing the population relative abundances, p_{ijk}, by the sample ones, f_{ijk}, with the usual continuity conventions $0^{\alpha} = 0$, $0 \log 0 = 0$). In the same way, the mutual information between X and the factor Y in the sample could be evaluated by means of the analogue estimator of the population index, $\iota^{\alpha}(X; Y)$, and the mutual information between X and the factor Z in the sample could be evaluated by means of the analogue estimator of the population index, $\iota^{\alpha}(X; Z)$.

From now on, and for the sake of notation simplicity, we will remove the value α to consider a generic population mutual information index, $I(X; \cdot)$, and the corresponding generic sample index, $\iota(X; \cdot)$.

3. PRELIMINARY RESULTS

As we have pointed out in the Introduction, we are first interested in testing the null hypothesis H_0: $I(X; Y) = I(X; Z)$ against the alternative hypothesis H_1: $I(X; Y) \neq I(X; Z)$, and in testing later the corresponding one-sided hypotheses. To define a test of those hypotheses it would be useful to examine the distribution of the statistic $\{[\iota(X; Y) - \iota(X; Z)] - [I(X; Y) - I(X; Z)]\}$.

In practice, when only small samples are available we could determine the exact values of the expectation and the variance of the preceding statistic for $\alpha = 2$ (that corresponds to the mutual information index associated with the Gini-Simpson diversity index), by following ideas in a previous paper (Gil et al. [13]). Then, on the basis of such values and by applying Chebyshev's inequality, we would construct a very conservative procedure for testing H_0 *vs* H_1.

In this paper we will assume large samples are available, so that we will obtain the asymptotic distribution of the statistic $\{[\iota(X; Y) - \iota(X; Z)] - [I(X; Y) - I(X; Z)]\}$ and we will construct an operative procedure for testing the considered hypotheses. This asymptotic distribution is now going to be presented in the following result

Theorem 3.1. The statistic $n^{1/2}\{[\iota(X; Y) - \iota(X; Z)] - [I(X; Y) - I(X; Z)]\}/\tau_n$ has an asymptotic standard normal distribution, as $n \to \infty$, whenever $\tau^2 > 0$, $\tau_n^2 > 0$, and where

$$\tau^2 = \left[\frac{\partial I}{\partial p}\right] \Sigma(p) \left[\frac{\partial I}{\partial p}\right]', \quad \Sigma(p) = \operatorname{diag}\{p_{111}, \ldots, p_{Mrs}\} - [p]'[p]$$

$$[p] = [p_{111} \ldots p_{Mrs}], \quad \left[\frac{\partial I}{\partial p}\right] = [V_{111} \ldots V_{Mrs}]$$

with

$$V_{ijk} = \alpha(\alpha - 1)^{-1} \Big[\Big(\sum_{m=1}^{s} p_{ijm}\Big)^{\alpha-1} - \Big(\sum_{h=1}^{M} \sum_{m=1}^{s} p_{hjm}\Big)^{\alpha-1} - \Big(\sum_{l=1}^{r} p_{ilk}\Big)^{\alpha-1} + \Big(\sum_{h=1}^{M} \sum_{l=1}^{r} p_{hlk}\Big)^{\alpha-1}\Big]$$

if $\alpha > 0$, $\alpha \neq 1$, and

$$V_{ijk} = \log \frac{\Big(\sum_{m=1}^{s} p_{ijm}\Big)\Big(\sum_{h=1}^{M} \sum_{l=1}^{r} p_{hlk}\Big)}{\Big(\sum_{h=1}^{M} \sum_{m=1}^{s} p_{hjm}\Big)\Big(\sum_{l=1}^{r} p_{ilk}\Big)}$$

if $\alpha = 1$, and τ_n^2 is the analogue estimate of τ^2. □

The asymptotic variance τ^2 would be equal to zero if an only if $p_{ijk}(V_{ijk} - c) = 0$, for all i, j, k, and c being a constant. In this particular situation, Theorem 3.1 does not make sense, but we can then use, for instance, an extension of Zvárová's result in [30]. This extension is gathered in the following result

Theorem 3.2. If $\tau^2 = 0$, then the statistic $2n\{[\iota(X; Y) - \iota(X; Z)] - [I(X; Y) - I(X; Z)]\}$ is asymptotically distributed as the quadratic form $Q_n = \lambda_1\xi_1^2 + \ldots + \lambda_{t-1}\xi_{t-1}^2$, where t is the number of $p_{ijk} > 0$, $\xi_1, \ldots, \xi_{t-1}$ are iid variables with the standard normal distribution, and $\lambda_1, \ldots, \lambda_{t-1}$ are the eigenvalues of $A\Xi$ (where A is the Hessian matrix, $[\partial^2 I/\partial p^2]$, and Ξ is the variance-covariance matrix of $t - 1$ idd variables $f_{ijk} - p_{ijk}$ for which $p_{ijk} > 0$).

4. PROCEDURES TO COMPARE THE SUITABILITY OF TWO FACTORS FOR STRATIFICATION

According to the notations in Sections 2 and 3, and in virtue of Theorem 3.1, we next detail the steps required:

(i) to test the null hypothesis H_0: $I(X; Y) = I(X; Z)$ against the alternative hypothesis H_1: $I(X; Y) \neq I(X; Z)$, at a specified significance level $\alpha \in (0, 1)$, on the basis of a large sample of size n from the community,

Step 1: Compute the sample values $\iota(X; Y)$ and $\iota(X; Z)$ (see Section 2).

Step 2: Compute the sample value τ_n, where

$$(\tau_n)^2 = \left[\frac{\partial\iota}{\partial f}\right] \Sigma(f) \left[\frac{\partial\iota}{\partial f}\right]', \quad \Sigma(f) = \operatorname{diag}\{f_{111}, \ldots, f_{Mrs}\} - [f]'\,[f]$$

$$[f] = f_{111} \ldots f_{Mrs}], \quad \left[\frac{\partial\iota}{\partial f}\right] = v_{111} \ldots v_{Mrs}]$$

with

$$v_{ijk} = -\alpha(\alpha - 1)^{-1}\Big[\Big(\sum_{m=1}^{s} f_{ijm}\Big)^{\alpha-1} - \Big(\sum_{h=1}^{M}\sum_{m=1}^{s} f_{hjm}\Big)^{\alpha-1} - \Big(\sum_{l=1}^{r} f_{ilk}\Big)^{\alpha-1} + \Big(\sum_{h=1}^{M}\sum_{l=1}^{r} f_{hlk}\Big)^{\alpha-1}\Big]$$

if $\alpha > 0, \alpha \neq 1$, and

$$v_{ijk} = \log \frac{\Big(\sum_{m=1}^{s} f_{ijm}\Big)\Big(\sum_{h=1}^{M}\sum_{l=1}^{r} f_{hlk}\Big)}{\Big(\sum_{h=1}^{M}\sum_{m=1}^{s} f_{hjm}\Big)\Big(\sum_{l=1}^{r} f_{ilk}\Big)}$$

if $\alpha = 1$.

Step 3: Compute the value

$$\lambda = \frac{n^{1/2}\iota[(X; Y) - \iota(X; Z)]}{\tau_n}$$

whenever $\tau_n > 0$.

Step 4: If $|\lambda| > \lambda_{\alpha/2}$ (where $\lambda_{\alpha/2}$ is the $\alpha/2$-percentage point of the standard normal distribution), then the null hypothesis H_0 must be rejected at the significance level α, otherwise it could be accepted.

(ii) to test the null hypothesis H_0': $I(X;Y) \geqq I(X;Z)$ against the alternative hypothesis H_1': $I(X;Y) < I(X;Z)$, at a specified significance level $\alpha \in (0,1)$, on the basis of a large sample of size n from the community,

Step 1, *Step* 2, and *Step* 3, as above

Step 4: If $\lambda < -\lambda_\alpha$, then the null hypothesis H_0' must be rejected at the significance level α, otherwise it could be accepted.

(iii) to test the null hypothesis H_0'': $I(X, Y) \leqq I(X;Z)$ against the alternative hypothesis H_1'': $I(X;Y) > I(X;Z)$, at a specified significance level $\alpha \in (0, 1)$, on the basis of a large sample of size n from the community

Step 1, *Step* 2, and *Step* 3, as above.

Step 4: If $\lambda > \lambda_\alpha$, then the null hypothesis H_0'' must be rejected at the significance level α, otherwise it could be accepted.

In all the previous cases, in addition to perform a test at a predetermined significance level α, it is a good statistical practice to record the p-value (or significance probability) that is, the smallest significance level for which the observed sample leads to the rejection of the null hypothesis. Thus, for our particular sample the p-value is given by: (i) (H_0 *vs* H_1) $p = 2[1 - \Phi(|\lambda|)]$; (ii) ($H_0'$ *vs* H_1') $p = \Phi(\lambda)$; (iii) (H_0'' *vs* H_1'') $p = 1 - \Phi(\lambda)$, where Φ denotes the cumulative standardized normal distribution function.

5. ILLUSTRATIVE EXAMPLE

We are now going to apply the preceding procedures to an ecological example.

The "ecological profile" of abundances of the species E for the factor $L$is established from the presence or absence of this species in the classes of the factor. The "mutual information between the species and the factor" is determined by the quantity of information brought up by the presence or the absence of the species for the factor. For each ecological factor, the level of the mutual information of the species and the study of their ecological profile allows one to obtain the "indicator value" of each species with respect to the classes of the factor. In addition, that level could further be used to examine the suitability of L for stratification, and to compare the suitability of two different factors, L_1 and L_2, for stratification, in estimating the diversity associated with the presence and the absence of E.

The suggested comparison is now illustrated by means of a practical example (Daget & Godron [5]) in which we consider the species *Myosotis scorpioides* (E) and the factors "global coverage of woody low strata" (L_1) and "stationary humidity"

(L_2), for which a sample of $n = 65$ sampling units drawn at random from the Liptov valley has supplied the data collected in Table 1 (where the strata for the factor L_1 are given by L_{11} = closed, L_{12} = rather closed, L_{13} = rather open, L_{14} = open, L_{15} = very open, L_{16} = extremely open, L_{17} = completely open, and the strata for the factor L_2 are given by L_{21} = extremely dry, L_{22} = very dry, L_{23} = dry, L_{24} = rather dry, L_{25} = median, L_{26} = rather humid, L_{27} = humid, L_{28} = very humid, L_{29} = extremely humid.

Table 1. Ecological Profiles of absolute abundances of the species *Myosotis scorpioides* for the factors Global Coverage of Woody Low Strata and Station Humidity. In each pair (a, b), the value a represents the number of sampling units in which the species is present and b represents the number of sampling units in which the species is absent.

Global coverage of woody low strata	Station Humidity								
	L_{21}	L_{22}	L_{23}	L_{24}	L_{25}	L_{26}	L_{27}	L_{28}	L_{29}
L_{11}	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
L_{12}	(0, 0)	(0, 0)	(0, 0)	(0, 1)	(0, 1)	(0, 1)	(0, 0)	(0, 0)	(0, 0)
L_{13}	(0, 0)	(0, 0)	(0, 1)	(0, 0)	(0, 2)	(0, 0)	(0, 0)	(1, 0)	(0, 0)
L_{14}	(0, 0)	(0, 0)	(0, 2)	(0, 4)	(0, 3)	(0, 0)	(0, 1)	(0, 0)	(0, 1)
L_{15}	(0, 0)	(0, 0)	(0, 2)	(0, 4)	(0, 4)	(1, 0)	(0, 0)	(0, 0)	(0, 0)
L_{16}	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 0)	(2, 0)	(0, 0)	(5, 0)	(1, 0)
L_{17}	(0, 0)	(0, 1)	(0, 3)	(1, 2)	(0, 5)	(1, 7)	(1, 0)	(0, 1)	(1,1)

As the community contains a large number of sampling units, we can assume that the considered sampling is random sampling with replacement. On the other hand, the size of the previous sample allows us to apply the procedure introduced in Section 4 to compare the goodness of L_1 and L_2 for stratification. According to the notations in that section, if we consider Shannon's index of mutual information we obtain $\lambda = -1{\cdot}4077$, so that the p-value in testing H_0: $I(E; L_1) = I(E; L_2)$ against the alternative hypothesis H_1: $I(E; L_1) \neq I(E; L_2)$ is equal to $p = 0{\cdot}1585$, and hence for all significance level $\alpha < 0{\cdot}1585$ the two factors are not significatively different with respect to the suitability for stratification of the community in estimating the diversity associated with the presence and the absence of E.

6. CONCLUDING REMARKS

The procedures we have just developed have been constructed by assuming that our inferences are based on a sample from the community. Similar procedures could be also defined by assuming that inferences would be based on two independent samples from the community. Those procedures would require greater sample size than that in this paper and would be defined by considering results such as those in Gil [11]. In addition to comments in Section 3 about the case in which only small samples are available, it would be interesting to develop studies based on the ideas in Smith & Grassle [27].

On the other hand, the study in the present paper could be complemented with the development of a statistical procedure to test the suitability of a factor for stratification. Thus, after concluding whether one factor is more suitable than another one, one could analyze the "degree of suitability" of the best factor by comparing the corresponding mutual information with its maximum value: the diversity of species in the community.

Finally, it would be also useful to discuss by means of simulated examples the following questions: (i) the goodness of the approximation in Theorem 3.1, in terms of the sample size n and the mutual information index, (ii) the gain in precision in estimating diversity, in terms of the corresponding mutual information associated with the classification process and the factor for stratification, according to different allocations.

APPENDIX

Proof of Theorem 3.1.

Indeed, whatever $\alpha > 0$ may be, the application of the first order Taylor expansion, or the Multivariate δ Method (cf., Bickel and Doksum, [4], Bishop et al. [6]) to the $M \times r \times s$-dimensional vector parameter $\boldsymbol{p} = (p_{111}, \ldots, p_{Mrs})$ and the $M \times r \times s$-dimensional vector value $\boldsymbol{f} = (f_{111}, \ldots, f_{Mrs})$, determines that

$$n^{1/2}\{[\iota(X; Y) - \iota(X; Z)] - [I(X; Y) - I(X; Z)]\} =$$
$$= n^{1/2}\left[\frac{\partial I}{\partial \boldsymbol{p}}\right][\boldsymbol{f} - \boldsymbol{p}]' + n^{1/2}R_n$$

(where R_n is the Lagrange remainder term) and, hence, the asymptotic distribution of $n^{1/2}\{[\iota(X; Y) - \iota(X; Z)] - [I(X; Y) - I(X; Z)]\}$, as $n \to \infty$, is $\mathcal{N}(0, \tau^2)$.

In addition, as f_{ijk} converges in probability to p_{ijk} as $n \to \infty$, then the analogue estimator of τ, τ_n, converges in probability to τ, and hence $n^{1/2}\{[\iota(X; Y) - \iota(X; Z)] - [I(X; Y) - I(X; Z)]\}/\tau_n$ has an asymptotic standard normal distribution, as $n \to \infty$, whenever $\tau > 0$ and $\tau_n > 0$. ☐

Proof of Theorem 3.2.

Following the ideas in [4], [28] and [30], we can now use the second order Taylor expansion of $2n[\iota(X; Y) - \iota(X; Z)]$ in a neighborhood of $\boldsymbol{p}$, so that

$$2n\{[\iota(X; Y) - \iota(X; Z)] - [I(X; Y) - I(X; Z)]\} =$$
$$= 2n\left[\frac{\partial I}{\partial \boldsymbol{p}}\right][\boldsymbol{f} - \boldsymbol{p}]' + n[\boldsymbol{f} - \boldsymbol{p}]\left[\frac{\partial^2 I}{\partial \boldsymbol{p}^2}\right][\boldsymbol{f} - \boldsymbol{p}]' + 2nU_n$$

(where U_n is the Lagrange remainder term), $[\partial^2 I/\partial \boldsymbol{p}^2]$ being the associated Hessian matrix.

This expansion, and the convergence in probability of $2n\{[\iota(X;Y) - \iota(X;Z)] - [I(X;Y) - I(X;Z)] + U_n\}$ to 0 guarantees that $2n\{[\iota(X;Y) - \iota(X;Z)] - [I(X;Y) - I(X;Z)]\}$ are asymptotically distributed as the statistic

$$Q_n = n[\boldsymbol{f} - \boldsymbol{p}]\left[\frac{\partial^2 I}{\partial \boldsymbol{p}^2}\right][\boldsymbol{f} - \boldsymbol{p}]'$$

which is a quadratic form of the variables $f_{ijk} - p_{ijk}$. As $p_{ijk} = 0$ implies that $f_{ijk} - p_{ijk} = 0$, then we can transform Q_n into a quadratic form that is asymptotically distributed, as $n \to \infty$, as the quadratic form of $t - 1$ iid variables $f_{ijk} - p_{ijk}$ for which $p_{ijk} > 0$ (where t is the number of $p_{ijk} > 0$).

Following ideas in [5] and [30], there exists a non-singular transformation so that $Q_n = \lambda_1 \xi_1^2 + \ldots + \lambda_{t-1}\xi_{t-1}^2$, where $\xi_1, \ldots, \xi_{t-1}$ are iid variables with the standard normal distribution, and $\lambda_1, \ldots, \lambda_{t-1}$ are the eigenvalues of $\boldsymbol{A\Xi}$ (with $\boldsymbol{A} = [\partial^2 I/\partial \boldsymbol{p}^2]$, and $\boldsymbol{\Xi}$ the variance-covariance matrix of $t - 1$ iid variables $f_{ijk} - p_{ijk}$ for which $p_{ijk} > 0$). ☐

ACKNOWLEDGEMENT

This research is supported in part by DGICYT Grant PS89-0169 NASA Grant NCC 2-275, and NSF PYI Grant DMC-84511622. Their financial support is gratefully acknowledged.

(Received December 22, 1989.)

REFERENCES

[1] G. P. Basharin: On a statistical estimate for the entropy of a sequence of independent random variables. Theory Probab. Appl. *4* (1959), 333—336.

[2] T. N. Bhargava and P. H. Doyle: A geometric study of diversity. J. Theoret. Biology *43* (1974), 241—251.

[3] T. N. Bhargava and V. R. R. Uppuluri: Sampling distribution of Gini's index of diversity. Appl. Math. Comput. *3* (1977), 1—24.

[4] P. J. Bickel and K. A. Doksum: Mathematical Statistics. Holden-Day, Inc., Oakland 1977.

[5] Z. W. Birnbaum: Introduction to Probability and Mathematical Statistics. Harper & Row, New York 1964.

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

[7] Ph. Daget et M. Godron: Analyse Fréquentielle de L'Écologie. Masson, Paris 1982.

[8] Z. Daróczy: Generalized information functions. Inform. and Control *16* (1970), 36—51.

[9] H. Emptoz: Informations Utiles et Pseudoquestionnaires. Thèse. Université Claude-Bernard, Lyon 1976.

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

[11] M. A. Gil: A note on the choice of the sample size in estimating the mutual information. Appl. Math. Comput. *30* (1989), 3, 125—132.

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

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

[14] I. J. Good: The population frequencies of species and the estimation of population parameters. Biometrika *40* (1953), 237—264.

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

[16] P. Holgate: The statistical entropy of a sample from a community of species. Biometrics *37* (1981), 795—799.

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

[18] J. A. Ludwig and J. F. Reynolds: Statistical Ecology. Wiley Interscience, New York 1988.

[19] D. R. Margalef: Information theory in ecology. General Systems *3* (1958), 36—71.

[20] T. K. Nayak: On diversity measures based on entropy functions. Comm. Statist. — Theory Methods *14* (1985), 1, 203—215.

[21] G. P. Patil and C. Taile: Diversity as a concept and its measurement. J. Amer. Statist. Assoc. *77* (1982), 548—567.

[22] E. C. Pielou: Ecological Diversity. Wiley Interscience, New York 1975.

[23] C. R. Rao: Diversity: its measurement, decomposition, apportionment and analysis. Sankhya Ser. A *44* (1982), 1, 1—22.

[24] A. Rényi: On measures of entropy and information. Berkeley Symp. Math. Statist. and Prob. *1* (1961), 547—561, Univ. California Press, Berkeley, Calif. 196.

[25] R. D. Routledge: Diversity indices. Which ones are admissible?. J. Theoret. Biology *76* (1979), 503—515.

[26] C. E. Shannon: A mathematical theory of communications. Bell System Technology Journal *27* (1948), 379—423, 623—656.

[27] W. Smith and F. Grassle: Sampling properties of a family of diversity measures. Biometrics *33* (1977), 283—292.

[28] K. Zografos, K. Ferentinos and T. Papaioannou: ϕ-divergence statistics: sampling properties and mutinomial goodness of fit and divergence tests. Comm. Statist. - Theory Methods 19 (1990), 1785—1802.

[29] J. Zvárová: The asymptotic distributions of sample information measure of dependence. Kybernetika *5* (1969), 1, 50—59.

[30] J. Zvárová: On asymptotic behaviour of a sample estimator of Rényi's information of order α. In: Trans. Sixth Prague Conf. Inform. Theory, Statist. Dec. Funct., Random Proc., Academia, Prague 1973, pp. 919—924.

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