

Copyright Notice

This electronic reprint is provided by the author(s) to be consulted by fellow scientists. It is not to be used for any purpose other than private study, scholarship, or research.

Further reproduction or distribution of this reprint is restricted by copyright laws. If in doubt about fair use of reprints for research purposes, the user should review the copyright notice contained in the original journal from which this electronic reprint was made.

On the measurement of association between plant species and environmental variables

Exequiel Ezcurra and Carlos Montaña

Instituto de Ecología, Apartado Postal 18-845, 11800 México, D. F.

SUMMARY

Two aspects can be recognized in the analysis of species response to environmental variation: (a) the degree of association between species and environmental factors and (b) the shape of the functional response.

The first aspect has been approached both from the standpoint of traditional statistics and from that of information theory. In this paper we analyze the use of log-likelihood ratios as a suitable form of measuring environmental correlation which can be used for data in the form of contingency tables. These ratios can be considered as a particular form of information measurements which have two very important properties: they are additive over different classes and have an asymptotic probabilistic distribution. They can, therefore, be used to test hypotheses about the distribution of species in relation to environmental variation. The departure from an expected distribution can be calculated for the whole set of data or for any particular factor class. Similarly, these statistics can be used to test hypotheses on niche breadth and resource overlap between species.

Information-theory measurements have been used widely for ecological analyses. Log-likelihood ratios combine the flexibility of use that is characteristic of information measurements with the rigour of statistical tests of hypothesis.

KEY-WORDS: *Log-likelihood ratios - Information theory - Contingency tables.*

RÉSUMÉ

On peut distinguer deux aspects dans l'analyse de la réponse des espèces aux variations de l'environnement : (a) la mesure d'association entre les espèces et les facteurs et (b) la forme de la fonction de réponse.

Le premier point a été abordé par les méthodes statistiques traditionnelles et par celles dérivées de la théorie de l'information. Dans ce travail on analyse l'utilisation des logarithmes des rapports de vraisemblance pour mesurer la corrélation entre espèces et facteurs, dans le cas où les données sont présentées sous la forme de tables de contingence. Ces rapports peuvent être considérés comme une forme particulière de mesures d'information qui présentent deux propriétés très importantes : ils peuvent s'additionner pour différentes classes et ils ont une distribution probabiliste asymptotique. Ils peuvent donc être utilisés pour tester des hypothèses relatives à la distribution des espèces en fonction des variations du milieu. L'éloignement des données à la distribution théorique attendue peut être calculé pour l'ensemble des données ou pour chaque classe de facteur. De la même façon, on peut utiliser cette statistique pour tester des hypothèses sur l'amplitude et le recouvrement des niches écologiques entre espèces.

Les mesures basées sur la théorie de l'information ont été largement utilisées dans les analyses écologiques. Les logarithmes des rapports de vraisemblance combinent la flexibilité d'utilisation qui est caractéristique des mesures d'information à la rigueur des tests statistiques.

MOTS-CLÉS : *Logarithme des rapports de vraisemblance - Théorie de l'information - Table de contingence.*

INTRODUCTION

Two aspects can be distinguished when analyzing the problem of species response to environmental variation: (a) the intensity or degree of association between species and environmental factors and (b) the shape of the functional response. The first aspect has been approached by some authors (*e. g.* : GREIG-SMITH, 1964; GOUNOT, 1969; GOLDSMITH & HARRISON, 1976) from the standpoint of traditional statistics (χ^2 tests, Anova's, etc.), while other authors have used information-theory statistics to measure environmental association (*e. g.* : GODRON, 1968; GUILLERM, 1971). The second aspect has been evaluated by fitting response curves either subjectively or quantitatively (*e. g.* : WHITTAKER, 1967; AUSTIN, 1980; AUSTIN & AUSTIN, 1980).

In this paper we analyze the first aspect of the problem, discussing the use of log-likelihood ratios as a suitable measurement of environmental correlation. These ratios have a mathematical structure that is similar to information-theory measurements, but they also have an associated probabilistic distribution. Thus, they combine the flexibility of use that is characteristic of information-theory measurements with the rigour of statistical tests of hypothesis.

THE STATISTICAL MEASUREMENT OF ASSOCIATION

The measurement of association between a species and an environmental variable implies measuring the relative abundance of the species for different values of the variable, in order to detect ranges of the variable for which the species shows higher—or lower—abundances than would be expected by chance. An ideal way to do this is to divide the variable in different classes, and count the species frequency for each class. This makes the problem particularly suitable for contingency table analysis.

Consider a species, distributed over s classes of an environmental factor such that f_{pi} indicates the relative frequency of the species in class i of the environmental factor. Let us assume that

$$f_{pi} = n_{pi}/n_{.i}$$

where n_{pi} is the absolute frequency of the species in class i (*i. e.* the number of quadrats in class i where the species was present), and $n_{.i}$ is the total number of quadrats in that class, such that $\sum_{i=1}^s n_{.i} = n$. Let us also assume that n_{ai} is the number of quadrats in class i where the species was absent. Thus,

$$n_{p.} = \sum_{i=1}^s n_{pi} \quad \text{and} \quad n_{a.} = \sum_{i=1}^s n_{ai};$$

and obviously the relative frequency of absences in class i is

$$f_{ai} = n_{ai}/n_{.i} = (n_{.i} - n_{pi})/n_{.i} = 1 - f_{pi}.$$

This notation can be summarized in the form of a two way contingency table.

	Classes of the environmental factor						
	1	2	...	<i>i</i>	...	<i>s</i>	Total
Freq. of presences	n_{p1}	n_{p2}	...	n_{pi}	...	n_{ps}	n_p
Freq. of absences	n_{a1}	n_{a2}	...	n_{ai}	...	n_{as}	n_a
Total	$n_{.1}$	$n_{.2}$...	$n_{.i}$...	$n_{.s}$	n

Two hypotheses on the distribution of the species along the gradient can be postulated:

H_0 . — The species is equidistributed along the gradient, and all the observed departures from equidistribution are due to random effects.

H_1 . — The species is *not* equidistributed along the gradient, and therefore the observed departures are due to the ecological response of the species to environmental variation or competition.

That is, under H_0 we suppose that the population relative frequencies (q_{pi} , also defined as the possibility of finding the species in a quadrat for the environmental class i) are equal to n_p/n for all i . Under H_1 we are assuming that population relative frequencies are not equal for all i , and are best estimated by sample frequencies (f_{pi}). In statistical terms the likelihood of the null hypothesis can be calculated as the product of the binomial probability for presences and absences over all the s classes

$$L(H_0) = \prod_{i=1}^s \frac{n_{.i}!}{n_{pi}! n_{ai}!} q_{pi}^{n_{pi}} q_{ai}^{n_{ai}} \tag{1}$$

Similarly, the likelihood of the alternative hypothesis is

$$L(H_1) = \prod_{i=1}^s \frac{n_{.i}!}{n_{pi}! n_{ai}!} f_{pi}^{n_{pi}} f_{ai}^{n_{ai}} \tag{2}$$

The likelihood ratio $\lambda(H_1, H_0)$ (LINDGREN, 1976, p. 246) is then

$$\Lambda = \lambda(H_1, H_0) = \frac{L(H_1)}{L(H_0)} \tag{3}$$

which can be simplified to

$$\Lambda = \frac{\prod_{i=1}^s f_{pi}^{n_{pi}} \prod_{i=1}^s f_{ai}^{n_{ai}}}{\prod_{i=1}^s q_{pi}^{n_{pi}} \prod_{i=1}^s q_{ai}^{n_{ai}}} \tag{4}$$

Now, it is a well-known fact that $2 \ln \Lambda$ is distributed as χ^2 (UPTON, 1978, p. 8-10; EVERITT, 1977, p. 79), the degrees of freedom being the number of restrictions on the data required to define the null hypothesis. Heuristic proofs of this can be found in KULLBACK *et al.* (1962) and in PIELOU (1975, p. 77). A more rigorous proof is presented by SILVEY (1970, p. 112-115). This relation is sometimes known as the *G* statistic (SOKAL & ROHLF, 1969) or as the "minimum discrimination information statistic" (m. d. i. s.; KULLBACK *et al.*, 1962). In general, given an observed absolute-frequency distribution (o_i) and a distribution expected under a null hypothesis (e_i), the *likelihood ratio test* (or *log-likelihood function*) is:

$$G = 2 \ln \Lambda$$

or,

$$G = 2 \sum_{i=1}^s o_i \ln \left(\frac{o_i}{e_i} \right)$$

Its large-sample distribution is similar to Pearson's X^2 test

$$X^2 = \sum_{i=1}^s \frac{(o_i - e_i)^2}{e_i}$$

with the same number of degrees of freedom.

For the particular case of the analysis of species frequencies, the *G* statistic becomes:

$$G = 2 \sum_{i=1}^s \left[n_{pi} \ln \frac{f_{pi}}{q_{pi}} + n_{ai} \ln \frac{f_{ai}}{q_{ai}} \right] \quad (5)$$

One of the most important properties of log-likelihood functions is their additivity over different data sets (KULLBACK *et al.*, 1962; LINDGREN, 1976). Hence, for our species-environmental factor problem, the equation can be written as

$$G = 2 \sum_{i=1}^s n_{pi} \ln \frac{f_{pi}}{q_{pi}} + 2 \sum_{i=1}^s n_{ai} \ln \frac{f_{ai}}{q_{ai}} \quad (6)$$

$$G = G_{\text{pres}} + G_{\text{abs}} \quad (7)$$

where G_{pres} is the log-likelihood function for the observed and expected *presences*, and G_{abs} is the log-likelihood function for the observed and expected *absences*. With a similar reasoning the overall log-likelihood value can be partitioned over the *s* environmental classes

$$G = \sum_{i=1}^s G_i \quad (8)$$

where G_i is the log-likelihood function for the observed and expected values corresponding to the *i*th. class of the factor. Additivity is a very important property of these functions. It permits the partitioning of the overall *G* value into several additive

components, and the null hypothesis becomes a combination of several partial hypotheses of interest. Each component is in itself a G -value, asymptotically distributed as χ^2 . In this sense, the analysis of log-likelihood functions is similar to the analysis of variance.

The overall value of G can be compared against the theoretical χ^2 distribution for $(s - 1)$ degrees of freedom. Thus, the statistical significance of the association between a species and an environmental factor can be known. Similarly, partial values (G_i) can be compared to test which classes have a greater departure from the expected values and to know the significance of these departures; *i. e.* the analysis of partial values can indicate which classes are significantly preferred or avoided. In the case of presence-absence data, the degrees of freedom for partial values are equal to one.

Generalization

Equation (5) can be written as

$$G = 2 \sum_{i=1}^s \left[n_{pi} \ln \frac{n_{pi}n}{n_{.i}n_p} + n_{ai} \ln \frac{n_{ai}n}{n_{.i}n_a} \right] \tag{9}$$

which can be simplified to:

$$G = 2 \left[\sum_{i=1}^s (n_{pi} \ln n_{pi} + n_{ai} \ln n_{ai}) - (n_p \ln n_p + n_a \ln n_a) - \sum_{i=1}^s n_{.i} \ln n_{.i} + n \ln n \right] \tag{10}$$

Equation (10) provides a formula for short-cut computations which can be generalized to any two-way contingency table. Suppose we are now analyzing quantitative data of a species distributed over s classes of an environmental factor. If we divide the abundance-scale for the species into c classes or categories, we can now create a $c \times s$ contingency table to summarize our data:

		Factor classes					Totals	
		1	2	...	j	...		s
Abundances classes	1	n_{11}	n_{12}	...	n_{1j}	...	n_{1s}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2j}	...	n_{2s}	$n_{2.}$

	i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{is}	$n_{i.}$

	c	n_{c1}	n_{c2}	...	n_{cj}	...	n_{cs}	$n_{c.}$
Totals		$n_{.1}$	$n_{.2}$...	$n_{.j}$...	$n_{.s}$	n

where n_{ij} represents the number of quadrats in the factor class j which had the species

in the abundance class i . The same equidistribution hypothesis as with presence-absence data can be formulated, and the G -statistic becomes:

$$G = 2 \left[\sum_{j=1}^s \sum_{i=1}^c n_{ij} \ln \frac{n_{ji}n}{n_i n_j} \right] \quad (11)$$

for $(c-1)(s-1)$ degrees of freedom \odot (Although simple in theory, the contingency table analysis of abundance data can be not so adequate in practice. It is commonly difficult to get enough field data in order to have each cell in the table properly represented). Equation (11) can easily be calculated using the short-cut formula

$$G = 2 \left[\sum_{j=1}^s \sum_{i=1}^c n_{ij} \ln n_{ij} - \sum_{i=1}^c n_i \ln n_i - \sum_{j=1}^s n_j \ln n_j + n \ln n \right] \quad (12)$$

Short-cut computations are well explained in SOKAL & ROHLF (1969) for two-way and higher order contingency tables. In general, for any two-way contingency table

$$G = 2(\sum x \log x \text{ for the cell frequencies} - \sum x \log x \text{ for the row and column totals} + n \log n).$$

It is important to note that the χ^2 approximation is only valid when the calculations are made using natural logarithms.

Information theory measurements

Log-likelihood functions are similar in many aspects to information theory measurements. By definition the information content of an event is equal to minus the logarithm of its probability (ABRAMSON, 1963)

$$I(x) = -\ln p(x), \quad (13)$$

then, a generalized log-likelihood function

$$G = 2 \sum_i n_i \ln \frac{f_i}{q_i} \quad (14)$$

can be rewritten as

$$G = 2 \left[-\ln \prod_{i=1}^s q_i^{n_i} - \left(-\ln \prod_{i=1}^s p_i^{n_i} \right) \right] \quad (15)$$

or

$$G = 2(-\ln(L(H_0)) - (-\ln(L(H_1)))) \quad (16)$$

That is, the log-likelihood function is equal to twice the information content of the event under the null hypothesis minus the information content of the event under the alternative hypothesis:

$$G = 2(I(H_0) - I(H_1)) \quad (17)$$

So really in terms of information theory the G -statistic is an "information gain" in the sense used by RÉNYI (1976) and CLIFFORD & WILLIAMS (1976). The links between

log-likelihood functions and information theory have been clear to many authors for a long time. KULLMAN *et al.* (1962) paper is a good example. χ^2 approximations of information—theory statistics have been used by LANCE & WILLIAMS (1968) for the comparison of floristic tables, and by PIELOU (1975) for the analysis of spatial association between species. ORLOCI (1978) explains in detail their use as multivariate resemblance functions.

Yet, in spite of this, the G -statistic has seldom been applied to the measurement of association between species and environmental factors (STRAHLER, 1977, 1978). Information-theory measurements that greatly resemble log-likelihood functions have been used extensively by some authors (*e. g.*: GODRON, 1966, 1968; GUILLERM, 1971; DAGET *et al.*, 1972; LEPART & DEBUSSCHE, 1980). GODRON (1966) uses a “mutual information” (ABRAMSON, 1963) measure for presence-absence data based on the equation

$$I(l; e) = \sum_{i=1}^s \frac{n_{pi}}{n} \log_2 \frac{n_{pi}n}{n_p \cdot n_{.i}} + \sum_{i=1}^s \frac{n_{ai}}{n} \log_2 \frac{n_{ai}n}{n_a \cdot n_{.i}} \quad (18)$$

where $I(l; e)$ is the “mutual information” between species e and factor l . Comparing equations (18) and (5), it is clear that

$$G = 2 (\ln 2)nI(l; e). \quad (19)$$

That is, to convert a “mutual information” value into the equivalent G value, it has to be multiplied by the number of quadrats, by the natural logarithm of 2 to change the logarithms to base e , and by 2 to approximate the log-likelihood function to a χ^2 distribution. By using his formula for “mutual information”, GODRON obtains a mean value *per quadrat* (or *relevé*) of association between the species and the environmental factor. This approach is very interesting as it permits the comparison of values from samples with different number of quadrats, but obscures a point of statistical interest: a high value of “mutual information” for say, four quadrats has much less statistical significance and can be more easily attributed to random effects than the same value for say, 400 quadrats.

Similarly, LEPART & DEBUSSCHE (1980) use the formula

$$I(i//e) = \frac{n_{pi}}{n_{.i}} \log_2 \frac{n_{pi}n}{n_p \cdot n_{.i}} + \frac{n_{ai}}{n_{.i}} \log_2 \frac{n_{ai}n}{n_a \cdot n_{.i}} \quad (20)$$

to measure the association between state i of the environmental factor and species e . Using the additive properties of the G -statistic, and comparing equations (8) and (20), it is again clear that

$$G_i = 2 (\ln 2)n_{.i}I(i//e). \quad (21)$$

Lepart and Debussche’s measure is a mean value per quadrat for each state i of the factor. As with Godron’s “mutual information”, standarization by $n_{.i}$ might make the values subjectively more comparable, but does not allow the statistical testing of the values and destroys their additivity.

Example

An example of application, taken from MONTAÑA (1982), analyzes the relations between environmental variables and vegetation types. The study area is the Rio

Manso watershed in the Andean-Patagonian Region of Argentina, located at 71° W and 41° S. According to CABRERA'S (1971) phytogeographic classification, this region involves part of the Subantarctic Province (Subantarctic Domain of the Antarctic Region) and part of the High-Andean Province (Andean-Patagonian Domain of the Neotropical Region). The Subantarctic Province is represented in the watershed by the Deciduous Forest District (characterized by two deciduous species of *Nothofagus*: *N. antarctica* and *N. pumilio*, and by a conifer: *Austrocedrus chilensis*), and by the Valdivian District (characterized by the evergreen *Nothofagus dombeyi*). The High-Andean Province is represented by the High-Andean Austral District, which is formed by high-altitude vegetation.

Altitude in the study area varies from 800 to 3,554 m. In the centre of the watershed, at an altitude of 845 m, the average anual rainfall is around 1,846 mm (measured between 1972 and 1976) and the mean temperature is 6.8° C. The mean temperature of the coldest month (July) is of 0.7° C, while that of the warmest month (January) is of 12.7° C. No month is frost-free. Precipitations (both rain and snow) are concentrated in fall and winter. Seventy percent of the yearly total falls between April and September. Thorntwaite's water balance shows no water deficit for all months of the year (GALLOPIN, 1978).

TABLE I. — Frequency distributions and G values for the vegetation type *Nothofagus dombeyi*-*N. pumilio* mixed forest against two environmental variables (altitude and slope) in the Rio Manso watershed, Rio Negro, Argentina (Tabulated G values are for $P = 0.005$).

ALTITUDE							
Variable classes	1	2	3	4	5	6	Totals
Presence	292	258	88	7	0	0	645
Absence	477	455	788	706	415	201	3042
Totals	769	713	876	713	415	201	3687
\underline{G} (observed)	180,44	141,26	38,58	217,27	159,61	77,30	814,46
\underline{G} (tabulated)	7,88 for each class						16,75

SLOPE							
Variable classes	1	2	3	4	5	6	Totals
Presence	108	193	199	123	21	1	645
Absence	627	834	929	558	87	7	3042
Totals	735	1027	1128	681	108	8	3687
\underline{G} (observed)	4,17	1,18	0,02	0,15	0,28	0,15	5,95
\underline{G} (tabulated)	7,88 for each class						16,75

Landforms are of glacial, glacio-lacustrine, lacustrine, fluvial and denudational origin (RABASSA, 1979), and have developed mainly on igneous extrusive and igneous intrusive rocks, and to a lesser extent, on metamorphic rocks.

More than 60 % of the watershed shows soils which have developed on volcanic ash. The rest of the area is formed mainly by rock outcrops or alluvial soils.

On the topographic map at a scale of 1 : 50000, the area of the watershed (247 km²) was divided into a square grid, each cell of the grid corresponding to 250 × 250 m in the field. For each one of the resulting 3687 cells, the vegetation type was determined (RODRIGUEZ GARCIA *et al.*, 1978), while the value of some environmental variables was measured on the topographic map (GALLOPIN *et al.*, 1980). In this example we shall analyze the association between two environmental variables (altitude and slope) and two vegetation types (mixed forest of *Nothofagus dombeyi* and *N. pumilio*, and forest of *N. antarctica* with *Schinus patagonicus*).

The altitude variable was divided into six categories with the following lower limits for each category: 800 m, 1000 m, 1200 m, 1400 m, 1600 m and 1800 m. The slope variable was also divided into six categories with the following lower limits: 0°, 10°, 20°, 30°, 40° and 50°. Tables I and II show the absolute frequency

TABLE II. — Frequency distributions and G values for the vegetation type *Nothofagus antarctica* forest with *Schinus patagonicus* against two environmental variables (altitude and slope) in the Rio Manso watershed, Rio Negro, Argentina (Tabulated G values are for P = 0.005).

		ALTITUDE						
Variable classes	1	2	3	4	5	6	Totals	
Presence	197	34	0	0	0	0	231	
Absence	572	679	876	713	415	201	3456	
Totals	769	713	876	713	415	201	3687	
G(observed)	290,31	2,95	113,36	92,26	53,70	26,01	578,59	
G(tabulated)	7,88 for each class						16,75	

		SLOPE						
Variables classes	1	2	3	4	5	6	Totals	
Presence	93	104	31	3	0	0	231	
Absence	642	923	1097	678	108	8	3456	
Totals	735	1027	1128	681	108	8	3687	
G(observed)	40,11	22,21	29,72	65,82	13,98	1,04	172,88	
G(tabulated)	7,88 for each class						16,75	

distribution of the vegetation types for the considered variables, while figure 1 shows the observed and expected relative frequencies.

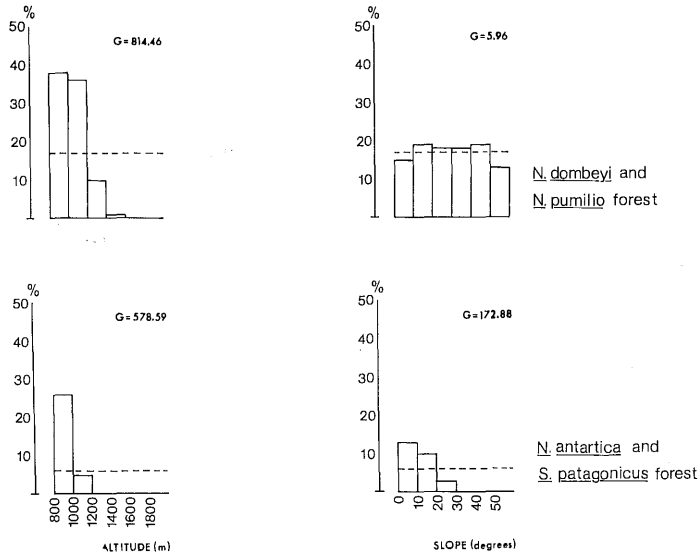


FIG. 1. — Observed and expected relative frequencies for two vegetation types (*Nothofagus dombeyi*-*N. pumilio* mixed forest, and *N. antarctica* forest with *Schinus patagonicus*) against two environmental variables (altitude and slope) in the Rio Manso watershed, Argentina (bars indicate observed frequencies, broken lines indicate expected frequencies).

The G test for the altitude variable shows that both vegetation types depart significantly from equidistribution ($P < 0.005$). Hence, we must conclude that both types of forest have marked altitudinal preferences within the study area. For the slope variable, on the other hand, only the *N. antarctica* forest shows a significant departure from equidistribution ($P < 0.005$), with a marked preference for low slopes and plains. The *N. dombeyi*-*N. pumilio* mixed forest does not show a significant preference for any particular slope class.

The analysis of partial G values (tables I and II) indicates that the *N. dombeyi*-*N. pumilio* mixed forest has a significantly high occurrence for the first two altitude classes, and a significantly low occurrence for the rest. The *N. antarctica* forest shows significant preference for the first altitude class only, and it is negatively associated with the last four classes. With respect to slope, it shows a significant preference for slopes of less than 20° , and a significant rejection for slopes between 20° and 50° . Although this type of forest does not appear in slopes of more than 50° , the scarcity of sites with this slope class does not permit to infer a statistically significant rejection for this category.

Lastly, it is important to note that in MONTAÑA'S (1982) original work, the sum of the G values for *all* the vegetation types in the study area was 9173 for altitude and 2846 for slope. This indicates that altitude is more important than slope in terms of

its capacity to explain the distribution of vegetation types. These conclusions were corroborated by a multiple discriminant analysis (MONTAÑA, 1982; MONTAÑA *et al.*, in prep.).

DISCUSSION AND CONCLUSIONS

There has been a convergence between traditional statistics in the form of log-likelihood ratios and the dictums of information theory. This is probably due to the computational advantages of dealing with the logarithmic expressions of probabilities rather than with probabilities themselves. The main virtues of these expressions are (a) their additivity and hence the possibility of adding or partitioning values in a way similar to variances, (b) their conversion into testable χ^2 approximations, and (c) their suitability for the analysis of contingency tables and for tabular analysis in general.

This last point is derived from the general properties of information-statistics, which deal with frequency distributions over discrete multistate classes. This is the case of contingency tables in general. Examples of the use of information-statistics in ecology can be seen in PIELOU (1972) and COLWELL & FUTUYMA (1971) for the measurement of niche breadth and species overlap; COLWELL (1974) and STEARNS (1981) for the analysis of periodic phenomena; LANCE & WILLIAMS (1968), ORLOCI (1968) and DALE *et al.* (1971) for the analysis of floristic tables. Of course, log-likelihood ratios can be used as suitable statistical tests for many of these applications, including tests of hypothesis on niche breadth and resource overlap between species.

In this context, analogies with thermodynamical entropy or with information (in the anthropic sense) have little ecological meaning; and therefore the measurement of association in bits and the use of \log_2 do not have a clear justification.

There is some similarity between the contingency table approach and direct ordination methods, which have been used extensively by WHITTAKER and his collaborators (WHITTAKER, 1956, 1960, 1967, 1978; GAUCH & WHITTAKER, 1972), and more recently by AUSTIN (1979, 1980). The contingency table approach implies the discretization of the environmental variable, and also of the abundance variable if continuous quantitative measurements of plant abundance are used. This implies necessarily the loss of some information and the definition (sometimes rather arbitrary) of frequency intervals. If these intervals are properly defined, the *G*-statistic can measure very efficiently the degree of association between a species and an environmental factor (but says little regards the form of the ecological response function).

Information-theory measurements have been used extensively for this purpose by some authors. Log-likelihood ratios can be considered as a particular form of information measurements which have two highly desirable properties: they are additive over different classes and have an asymptotic probability distribution. Thus, the significance of the departure from the expected values in any frequency distribution (of which the species-environment correlation problem is a particular case) can be partitioned over different factor classes. Both the overall significance for the whole contingency table and the significance for any particular class can be calculated.

ACKNOWLEDGMENTS

The authors thank Dr. P. GREIG-SMITH for the critical reading of a first draft of the manuscript and for helpful suggestions. Miguel EQUIHUA contributed to our ideas with stimulating discussions.

This work was done in the Instituto de Ecología, México. The authors want to thank the Mexican Council for Science and Technology (CONACYT); the Ford Foundation and the Fundación Bariloche for partial financial assistance.

REFERENCES

- ABRAMSON N., 1963. — *Information Theory and Coding*. McGraw-Hill, New York.
- AUSTIN M. P., 1979. — Current approaches to the non-linearity problem in vegetation analysis. In: G. P. PATIL & M. L. ROSENZWEIG, eds., *Contemporary Quantitative Ecology and Related Econometrics*, vol. 12, Statistical Ecology Series. International Co-operative Publishing House, Maryland.
- AUSTIN M. P., 1980. — Searching for a model for use in vegetation analysis. *Vegetatio*, **42**, 11-21.
- AUSTIN M. P. & AUSTIN B. O., 1980. — Behaviour of experimental plant communities along a nutrient gradient. *Journal of Ecology*, **68**, 891-918.
- CABRERA A. L., 1971. — Fitogeografía de la República Argentina. *Boletín Sociedad Argentina de Botánica*, **14**, 1-42.
- CLIFFORD H. T. & WILLIAMS W. T., 1976. — Similarity measures. In: W. T. WILLIAMS, ed., *Pattern Analysis in Agricultural Science*. CSIRO, Melbourne, and Elsevier, Amsterdam.
- COLWELL R. K., 1974. — Predictability, constancy, and contingency of periodic phenomena. *Ecology*, **55**, 1148-1153.
- COLWELL R. K. & FUTUYMA D. J., 1971. — On the measurement of niche breadth and overlap. *Ecology*, **52**, 567-576.
- DAGET P., GODRON M. & GUILLERM J. L., 1972. — Profils écologiques et information mutuelle entre espèces et facteurs écologiques. In: E. VAN DER MAAREL & R. TUXEN, eds., *Grundfragen und Methoden in der Pflanzensoziologie*. Junk, The Hague.
- DALE M. B., LANCE G. N. & ALBRECHT L., 1971. — Extensions to information analysis. *Computer Journal*, **3**, 29-34.
- EVERITT B. S., 1977. — *The Analysis of Contingency Tables*. Chapman & Hall, London.
- GALLOPIN G. C., 1978. — Estudio ecológico integrado de la cuenca del Río Manso Superior (Río Negro, Argentina). I. Descripción general de la cuenca. *Anales de Parques Nacionales*, **14**, 161-230.
- GALLOPIN G. C., BRANDANI A., GOMEZ I., GRIGERA D., GROSS M., RAMASSOTTO A., RODRIGUEZ GARCIA D. & TSUIJ S., 1980. — *Integrated Ecological Study of the Upper River Manso Basin (Río Negro, Argentina)*. IV. *Quantitative Topographical Analysis*. Fundación Bariloche, Argentina.
- GAUCH H. G. & WHITTAKER R. H., 1972. — Coenocline simulation. *Ecology*, **53**, 446-451.
- GODRON M., 1966. — Application de la théorie de l'information à l'étude de l'homogénéité et de la structure de la végétation. *Æcologia Plantarum*, **1**, 187-197.
- GODRON M., 1968. — Quelques applications de la notion de fréquence en écologie végétale. *Æcologia Plantarum*, **3**, 185-212.
- GOLDSMITH F. B. & HARRISON C. M., 1976. — Description and analysis of vegetation. In: S. B. CHAPMAN, ed., *Methods in Plant Ecology*. Blackwell Scientific Publications, Oxford.
- GOUNOT M., 1969. — *Méthodes d'Étude Quantitative de la Végétation*. Masson et Cie, Paris.
- GREIG SMITH P., 1964. — *Quantitative Plant Ecology*, 2nd. edn., Butterworths, London.
- GUILLERM J. L., 1971. — Calcul de l'information fournie par un profil écologique et valeur indicatrice des espèces. *Æcologia Plantarum*, **6**, 209-225.
- KULLBACK S., KUPPERMAN M. & KU H. H., 1962. — Tests for contingency tables and Markov chains. *Technometrics*, **4**, 573-608.
- LANCE G. N. & WILLIAMS W. T., 1968. — Note on a new information statistic classificatory program. *Computer Journal*, **11**, 195.
- LEPART J. & DEBUSSCHE M., 1980. — Information efficiency and regional constellation of environmental variables. *Vegetatio*, **42**, 85-91.
- LINDGREN B. W., 1976. — *Statistical Theory*, 3rd. edn., MacMillan, New York.
- MONTAÑA C., 1982. — *Las comunidades de ñire (Nothofagus antarctica (Forst.) Oerst.) de la cuenca del Río Manso Superior (Río Negro, Argentina)*. Tesis doctoral. Universidad Nacional de Córdoba, Argentina.

- ORLOCI L., 1968. — Information analysis in phytosociology: partition, classification and prediction. *Journal of Theoretical Biology*, **20**, 271-284.
- ORLOCI L., 1978. — *Multivariate Analysis in Vegetation Research*, 2nd edn., Junk, The Hague.
- PIELOU E. C., 1972. — Niche width and overlap: a method for measuring them. *Ecology*, **53**, 687-692.
- PIELOU E. C., 1975. — *Ecological Diversity*. John Wiley & Sons, New York.
- RABASSA J., 1979. — *Mapa Geomorfológico de la Cuenca del Rio Manso Superior, Provincia de Rio Negro, Argentina*. Universidad Nacional del Comahue. Publicaciones Facultad de Humanidades. 4. Neuquén, Argentina.
- RODRIGUEZ GARCIA D., SOURROUILLE A., GALLOPIN G. C. & MONTAÑA C., 1978. — Estudio ecológico integrado de la cuenca del Rio Manso Superior (Rio Negro, Argentina). II. Tipos de vegetación. *Anales de Parques Nacionales*, **14**, 231-250.
- RENYI A., 1976. — *Cálculo de Probabilidades*. Ed. Reverté, Barcelona.
- SILVEY S. D., 1970. — *Statistical Inference*. Penguin Books, Middlesex.
- SOKAL R. R. & ROHLF F. J., 1969. — *Biometry*. W. H. Freeman. San Francisco.
- STEARNS S. C., 1981. — On measuring fluctuating environments: predictability, constancy, and contingency. *Ecology*, **62**, 185-199.
- STRAHLER A. H., 1977. — Response of woody species to site factors in Maryland, USA: evaluation of sampling plans and of continuous and binary measurement techniques. *Vegetatio*, **35**, 1-19.
- STRAHLER A. H., 1978. — Binary discriminant analysis: a new method for investigating species-environment relationships. *Ecology*, **59**, 108-116.
- UPTON G. J. G., 1978. — *The Analysis of Cross-tabulated Data*. John Wiley & Sons, Chichester.
- WHITTAKER R. H., 1956. — Vegetation of the Great Smoky Mountains. *Ecological Monographs*, **26**, 1-80.
- WHITTAKER R. H., 1960. — Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, **30**, 279-338.
- WHITTAKER R. H., 1967. — Gradient analysis of vegetation. *Biological Review*, **42**, 207-264.
- WHITTAKER R. H., 1978. — Direct gradient analysis. In: R. H. WHITTAKER, ed., *Ordination of Plant Communities*. Junk, The Hague.

