FiskDir. Skr. Ser. HavUnders., 15: 234—245.

THE LOGARITHMIC SERIES AND THE LOGNORMAL DISTRIBUTION APPLIED TO BENTHIC INFAUNA FROM PUGET SOUND, WASHINGTON, U S A [1]

By

ULF LIE

Department of Oceanography, University of Washington,
Seattle, Wash., USA.

INTRODUCTION

Collections from natural communities of organisms are characterized by unequal abundances of the various species, and several mathematical models have been proposed to represent the distribution of specimens among the species (TAYLOR 1953, WILLIAMS 1964). Two of the most frequently used models are the logarithmic series (FISHER, CORBETT and WILLIAMS 1943) and the lognormal distribution (PRESTON 1948). The present paper is an attempt to test the validity of these models for communities of benthic organisms.

R. A. FISHER demonstrated that the negative binomial distribution could be applied to the frequency distribution of species abundances in samples from natural communities where the species were distributed randomly and the abundances of the various species were distributed in accordance with a Gamma-function. The proportion of species with abundance n in a sample is then expressed by:

$$P_{(n)} = \frac{(k + n - 1)!\, p^n}{(k - 1)!\, n!\, (1 + p)^{k+n}}$$

where p is a parameter related to the sample size and k is an inverse measure of the variability in the abundances of the species. When applied to a collection the zero term may be omitted, since no information is available about the number of species present in the community but not included in the sample. Furthermore, the range of species abundances is

[1] Contribution No. 505, Department of Oceanography, University of Washington.

Contribution given in honour of Gunnar Rollefsen at his 70th birthday.

normally so large that k becomes nearly zero. Thus, by omitting the zero term and setting $k = 0$, and by substituting α for $1/(k-1)!$ and x for $p/(1+p)$ in the formula for the negative binomial distribution, FISHER arrived at the logarithmic series:

$$P_{(n)} = \alpha x^n/n$$

where $P_{(n)}$ is the proportion of species in the sample with n specimens. The relationship between number of species (S) and number of specimens (N) in the logarithmic series is:

$$S = \alpha \ln\left(1 + \frac{N}{\alpha}\right)$$

Total number of specimens in replicate samples from a community satisfying the logarithmic series will be distributed according to the negative binomial distribution and total number of species according to a POISSON-series (QUENOUILLE 1949). KENDALL (1948) showed that the logarithmic series could originate from various types of population growth, and a comprehensive account of the application of the logarithmic series to various aspects of quantitative biology is given by WILLIAMS (1964).

PRESTON (1948) questioned two consequences of the logarithmic series model, namely, that species with few specimens in a community are more numerous than species with intermediate abundances, and that no upper limit to the number of species is defined. He stated that the number of species is limited, and that it is not reasonable to expect that a major part of the species are of very low abundance; in reality the majority of the species are of intermediate abundance. PRESTON showed that the distribution is normal when the abundances are grouped in geometric classes differing by a factor of two, i.e., the distribution is lognormal. However, if a collection is small compared to the total number of species present in the community the distribution is a truncated lognormal; in such situations many of the species appear to be rare and the data fit a logarithmic series. By doubling the sample size the mode of the lognormal distribution shifts to one higher class and the distribution thus becomes less and less truncated with increasing sample size.

WILLIAMS (1953) agreed that the lognormal distribution often is preferable to the logarithmic series; however, he found that classes differing by a factor of three are better, because by PRESTON's method half of the number of species with two specimens are included in the first class and the other half in the second class. By WILLIAMS' $\times 3$ classes the species in the first class have one specimen, in the second class two to four, in the third class five to thirteen, and in the fourth class fourteen to forty specimens, such that the midpoints of the classes are one, three, nine and twenty-seven specimens.

MATERIALS AND METHODS

The present material was collected quarterly (seven cruises) during 1963–64 at eight stations in Puget Sound in from five to ten replicate samples with a 0.1 m^2 van Veen grab (LIE 1968). All the polychaetes, crustaceans, lamellibranchs and echinoderms from the first two cruises (January–February and April–May 1963) were identified and counted; in the samples from August and November the polychaetes were identified only from stations 2 and 7; in the samples from 1964 (three cruises) polychaetes were not identified. Therefore, the logarithmic series has been applied only to the data from the two first cruises when the number of species was sufficiently high for meaningful plotting of the first ten classes of the series (species with from one to ten specimens).

The logarithmic series was fitted by finding x from table 146, page 308, in WILLIAMS (1964), and α was calculated from the formula:

$$\alpha = N(1 - x)/x$$

The logarithmic series can be fitted when x and α are known (page 235). The deviation of the logarithmic series from observed frequencies was estimated by the formula:

$$D = \left[\sum_{i=1}^{n} (\log A_i - \log B_i)^2 / n \right]^{1/2}$$

where A_i is the observed frequency, B_i the frequency expected from the logarithmic series, and n the number of classes.

The lognormal distribution was tested by plotting the cumulative percentage of the $\times 3$ classes on logarithmic probability paper; a log-normal distribution is represented by a straight line on such paper (AITCHISON and BROWN 1957).

The SHANNON–WIENER function of information theory (SHANNON and WEAVER 1963):

$$H = - \sum_{i=1}^{s} p_i \log_2 p_i$$

is a measure of the complexity of an animal community; the more similar the abundances of the various species and the higher the number of species, the higher the diversity of the community as measured by the

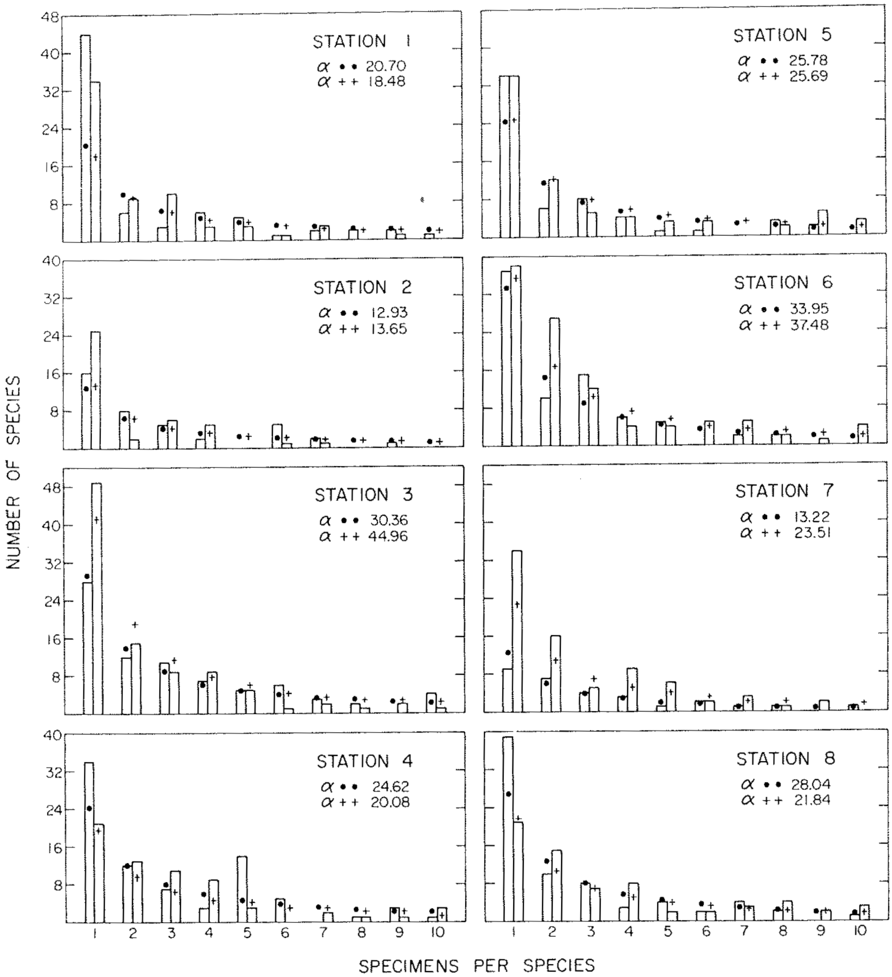

Fig. 1. Observed frequencies of specimens per species (bars) and frequencies expected from the logarithmic series in January–February (dots) and April–May (crosses) 1963 of benthic infauna from eight permanent stations in Puget Sound.

formula. The relative diversity (H_r) or "evenness" (PIELOU 1966) is the ratio between the observed diversity and the maximum diversity the community could contain, i.e., if all the species were equally abundant:

$$H_r = H/H_{max} \ (H_{max} = \log_2 S)$$

Thus the relative diversity represents only the distribution of specimens among the species and is unaffected by variation in the total number of species in a sample. SHANNON and WEAVER (1963) defined redundancy as one minus the relative information, and the redundancy for a community of organisms is therefore a measure of the level of dominance in the community. Redundancy expressed in percent by the formula:

$$R = 100(1 - H_r)$$

has therefore been used in the present paper as a measure of the dominance of the faunal assemblages in Puget Sound.

RESULTS AND DISCUSSION

THE LOGARITHMIC SERIES

Fig. 1 shows the observed frequencies and the frequencies expected from the logarithmic series of species with from one to ten specimens in the samples from the eight stations in Puget Sound during January–February and April–May 1963. There is generally a reasonably good fit, but in seven of the sixteen cases the discrepancy between observed and expected frequencies was larger than could be expected when tested by a Chi^2 test (Table 1). Since the logarithmic series were fitted by determining two parameters, x and α, the degrees of freedom in the Chi^2 tests equal the number of classes minus two.

Table 1. Chi^2 test of goodness of fit of observed frequencies to frequensies expected from the logarithmic series for benthic infauna from eight stations in Puget Sound.

Station	Date	Chi^2	Degrees of Freedom	p(0.05)
1	February 1963	37.8	5	11.07
1	April 1963	25.2	5	11.07
2	February 1963	2.8	3	7.81
2	May 1963	21.0	3	7.81
3	February 1963	3.7	6	12.59
3	April 1963	8.6	6	12.59
4	January 1963	25.7	5	11.07
4	April 1963	10.6	5	11.07
5	February 1963	11.5	5	11.07
5	April 1963	8.2	5	11.07
6	February 1963	8.6	5	11.07
6	April 1963	12.9	6	12.59
7	February 1963	2.2	3	7.81
7	April 1963	21.8	5	11.07
8	February 1963	9.8	5	11.07
8	April 1963	6.4	5	11.07

The discrepancy between the observed and the expected frequencies was particularly high at stations with a high degree of numerical dominance, and the deviation from the logarithmic series (D) for the first five classes (species with from one to five specimens) appeared to be positively correlated to the redundancy (R) (Fig. 2). The correlation is weak, probably largely because of sampling variability for the rare species, but KENDALL's rank correlation coefficient was 0.61 ($p < 0.01$). A test of the relationship between goodness of fit to the logarithmic series and redundancy could be performed if the redundancy could be manipulated and the subsequent change in the goodness of fit studied. Stations 1 and 3 were quite similar in species composition and about the same species were numerically dominant at the two stations (LIE 1968); however, station 1 had a very high and station 3 a very low redundancy. When the data from these two stations from February 1963 were pooled, the combined redundancy as expected became lower than at station 1 but higher than

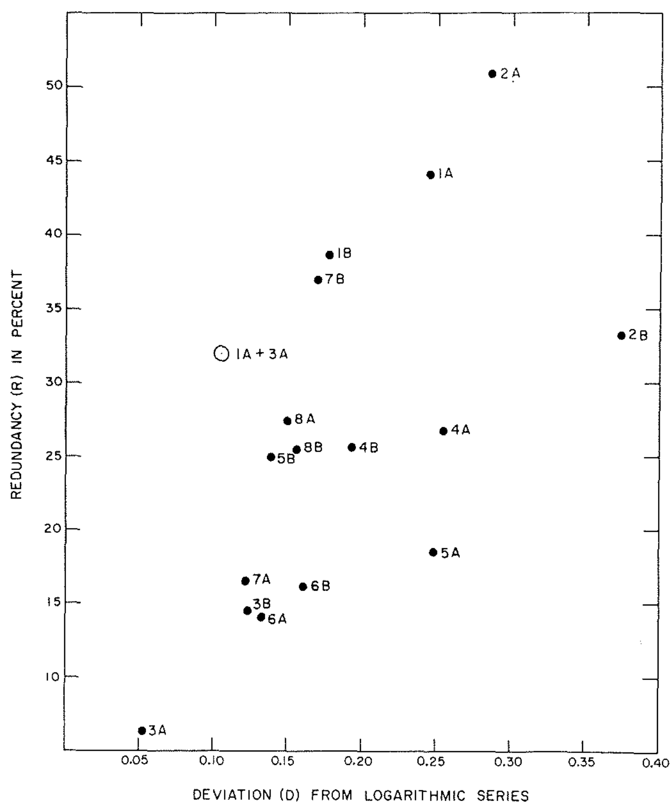

Fig. 2. The relationship between redundancy and the deviation of observed frequencies from frequencies expected from the logarithmic series at eight stations in January–February (A) and April–May (B) 1963.

at station 3. Consequently the deviation from the logarithmic series (Fig. 2) became lower than for station 1 but higher than for station 3.

Several examples of poor fit to the logarithmic series in collections of insects with strong numerical dominance by one or a few species are discussed by WILLIAMS (1964). A characteristic feature of these situations, as for the benthic infauna, was a too high number of species with only one specimen. In communities with strong dominance the basic assumption of a Gamma-distribution of the species abundances may not be satisfied.

One of the underlying assumptions for the logarithmic series is that the species are randomly and independently distributed. However, non-randomness of spatial distribution occurs frequently in nature (HAIRSTON 1959) and LIE (1968) showed that most of the numerically dominant species at the stations in Puget Sound had varying degrees of patchiness. The measured patchiness is in part a function of sampling design and sampling techniques. Certain sampling techniques, such as using light traps to attract and capture insects (WILLIAMS 1964) and trawling for bottom fish (TAYLOR 1953), have provided data with particularly good fit to the logarithmic series. However, these sampling techniques tend to obscure real patchiness of the fauna and thus they induce randomness. In the benthic communities in Puget Sound a considerable portion of species are commensal (viz. members of the polychaete family Polynoidae and the crustacean family Pinnotheridae), and consequently the requirement of independence is often not satisfied.

Benthic communities, particularly in the arctic and boreal regions, are characterized by dominance in abundance or biomass by a small number of species (THORSON 1957) and normally there is patchiness and a considerable degree of interdependence among the species. As the basic assumptions for the logarithmic series are randomness of spatial distribution and independence among the species, and as the series appears to be adversely affected by numerical dominance, it is doubtful if the series generally is a representative model for the relationship among species and specimens in a benthic infaunal community.

THE LOGNORMAL DISTRIBUTION

The data from January–February and April–May 1963 were grouped in $\times 3$ classes (WILLIAMS 1953) and the cumulative percentages were plotted on logarithmic probability paper (Fig. 3). The curves are reasonably straight, which indicates that the distributions are close to the lognormal. However, all the curves are characterized by having the median (50% probability) in the first or second class, indicating that half or more of the species had not been sampled. By increasing the sample

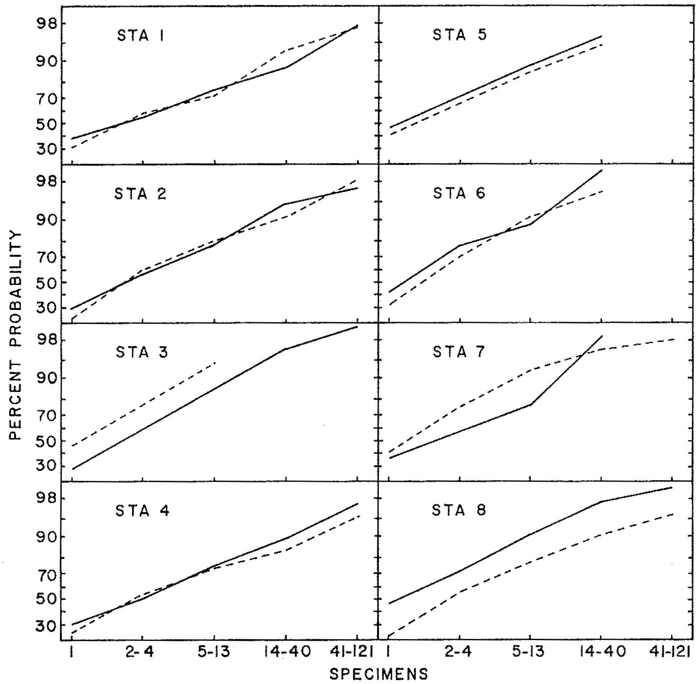

Fig. 3. Cumulative percentage distribution of species in ×3 classes of abundance. Solid line; January—February 1963; broken line: April—May 1963.

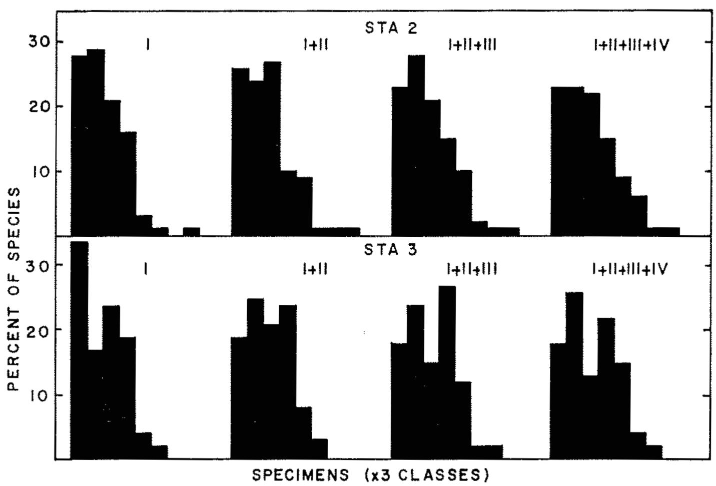

Fig. 4. Percentage frequency distribution of species in ×3 classes of abundance at stations 2 and 3. Combined data from January–February (I), April–May (II), August (III), and November (IV) 1963.

size the median should shift to the right on the abscissa (PRESTON 1948), i.e., the first class should contain a progressively lower percentage of the total number of species. WILLIAMS (1953) showed that the mode of the lognormal distribution of insects caught in light traps shifted to the right when the sampling time was increased from an eighth of a year to eight years. The faunal assemblages in Puget Sound exhibited small variations in number of species and the numerically dominant species remained largely the same throughout the investigated period (LIE 1968). Combining the data from different cruises for a station would then, as for WILLIAMS' insect data, in effect be increasing the sample size. Fig. 4 shows the percentage frequencies of the total fauna at station 2 and the non-polychaetes at station 3 when data were combined for the four cruises in 1963. A significant shift of the mode can hardly be ascertained; even after combining all four cruises the curves are strongly truncated, although a reduction in the percentages of the first class did take place.

The constancy of total number of species during the investigated period does not necessarily imply that the same species always were present. Fifteen of the species (23%) found at station 2 in May were not found in February, eleven species (17%) found in August were not found during the two preceding cruises, and nine species (12%) found in November were not found during the three preceding cruises. Similarly, twenty-five of the species (45%) found at station 3 in April were not found in February, eleven species (19%) found in August were not found during the two preceding cruises, and seven species (12%) were not found during the three preceding cruises. Therefore, by increasing the sample size it appears that the chance of encountering new species was nearly as high as the chance of adding new specimens to earlier recorded rare species. HAIRSTON and BYERS (1954) did not find a significant shift of the mode of the lognormal distribution with increasing sample size, and they argued that the reason for this was the extreme patchiness of the rare species.

When communities are considered more or less well defined stages in a "continuum" (WHITTAKER 1967), a certain number of the rare species will be representatives of neighbouring communities. Increasing the sample size will lead to discovery of more of these rare, transient species, and the lognormal distribution will always be strongly truncated. An upper limit of the number of species in a community cannot be defined (BRIAN 1953), except the limit set by the respective zoogeographical regime. LIE (1968) found a gradual transition in faunal composition among the eight Puget Sound stations. The benthic communities appear to be arrayed in a continuum that is directly related to physical properties of the substrate. When the data from all seven seasons were combined for each station and for all the stations, thus effectively increasing sample size, the lognormal distributions were still strongly truncated for the individual stations (Fig. 5). In contrast, the distribution for the combined data was considerably less truncated. Therefore, it appears that the sampling at the eight stations in Puget Sound during 1963–64 was sufficient to encompass the majority of the species in this relatively well defined geographical area.

Complete or nearly complete lognormal distributions have been demonstrated in large samples from well defined geographical areas such as islands (PRESTON 1962) and in samples covering a very large area or collected over a very long time (WILLIAMS 1964). Such samples will contain species from more than one community, and the lognormal distribution is therefore a model for the "between habitat" species diversity (MACARTHUR 1965).

Benthic communities may be identified as "statistical units" based on the presence of characteristic species (THORSON 1957) or ordinated in multi-dimensional space by multivariate analysis (CASSIE and MICHAEL 1968), but regardless of technique of analysis the communities must be considered "open" in the sense that there are no distinct borders between them. The lognormal distribution in such communities will be strongly truncated and therefore not be a better model for the distribution of specimens among the species than the logarithmic series.

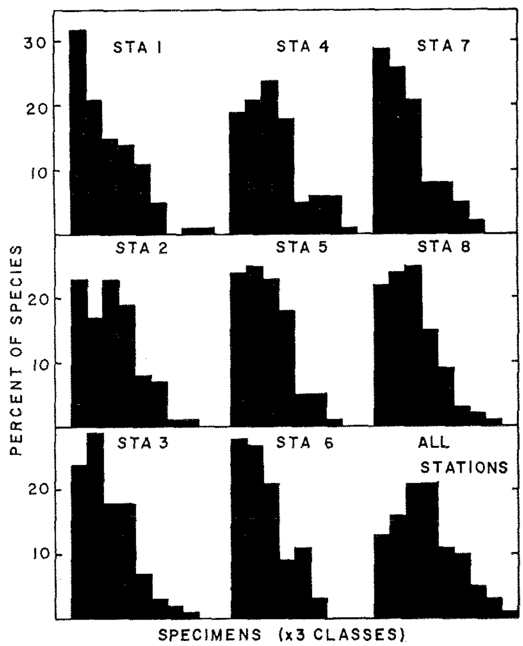

Fig. 5. Percentage frequency distribution of species in $\times 3$ classes of abundance. Combined data from 1963–1964 for single stations and for all stations.

SUMMARY

1. Benthic infauna collected at eight stations in Puget Sound during 1963–1964 was tested for agreement with the logarithmic series and the lognormal distribution.
2. The data agreed only moderately well with the logarithmic series; seven of sixteen cases showed statistically significant differences between observed frequencies and frequencies expected from the logarithmic series. Furthermore, there was an inverse relationship between goodness of fit to the logarithmic series and the level of numerical dominance of the faunal assemblages. This relationship, and the fact that the underlying assumptions for the logarithmic series are not applicable in benthic communities, makes the series a doubtful model for the relationship between species and specimens in a benthic community.
3. The lognormal distribution fitted the data reasonably well, but the curves were always strongly truncated. Increasing the sample size reduced the percentage of species in the first class somewhat, but only when the data from all the seasons and all the stations were combined was a significant part of the distribution "unveiled". The lognormal distribution therefore represents the geographical region reasonably well, but for each station the curves are too truncated to represent better models than the logarithmic series.

ACKNOWLEDGMENTS

This work was supported by U.S. PUBLIC HEALTH SERVICE GRANT GM-10817 from the Division of General Medical Sciences to K. Banse and U. Lie.

I wish to express my thanks to Dr. J. C. KELLEY who suggested the formula for estimating the deviation between theoretical and observed frequencies, and to Dr. B. W. FROST for criticisms and suggestions to the manuscript.

REFERENCES

AITCHISON, J. and BROWN, J. A. C. 1957. *The lognormal distribution with special reference to its uses in economics*. Univ. Cambridge Press, 176 pp.

BRIAN, M. V. 1963. Species frequencies in random samples from animal populations. *J. Anim. Ecol.*, *22*: 57–64.

CASSIE, R. M. and MICHAEL, A. D. 1968. Fauna and sediments of an intertidal mudflat: A multivariate analysis. *J. exp. mar. Biol. Ecol.*, *2*: 1–23.

FISHER, R. A., CORBETT, A. S. and WILLIAMS, C. B. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.*, *12*: 42–58.

HAIRSTON, N. G. 1959. Species abundance and community organization. *Ecology*, *40*: 404–416.

HAIRSTON, N. G. and BYERS, G. W. 1954. The soil arthropods of a field in southern Michigan. *Contr. Lab. vertebr. Biol. Univ. Mich.*, *64*: 1–37.

KENDALL, D. G. 1948. On some modes of population growth leading to R. A. FISHER's logarithmic series distribution. *Biometrika*, *35*: 6–15.

LIE, U. 1968. A quantitative study of benthic infauna in Puget Sound, Washington, U.S.A., in 1963–1964. *FiskDir. Skr. Ser. HavUnders.*, *14*: 229–556.

MACARTHUR, R. H. 1965. Patterns of species diversity. *Biol. Rev.*, *40*: 510–533.

PIELOU, E. C. 1966. Species-diversity and pattern-diversity in the study of ecological succession. *J. Theoret. Biol.*, *10*: 370–383.

PRESTON, F. W. 1948. The commonness, and rarity of species. *Ecology*, *29*: 254–283.

— 1962. The canonical distribution of commonness and rarity. *Ecology*, *43*: 185–215, 410–432.

QUENOUILLE, M. H. 1949. A relation between the logarithmic, POISSON, and negative binomial series. *Biometrics*, *5*: 162–164.

SHANNON, C. E. and WEAVER, W. 1963. *The mathematical theory of communication.* Univ. Illinois Press, Urbana, 117 pp.

TAYLOR, C. C. 1953. Nature of variability in trawl catches. *Fishery Bull. Fish Wildl. Serv. U.S.* *83*: 145–166.

THORSON, G. 1957. Bottom communities (sublittoral and shallow shelf). *Mem. geol. Soc. Am.*, *67*(1): 461–534.

WHITTAKER, R. H. 1967. Gradient analysis of vegetation. *Biol. Rev.*, *42*: 207–264.

WILLIAMS, C. B. 1953. The relative abundance of different species in a wild animal population. *J. Anim. Ecol.*, *22*: 14–31.

— 1964. *Patterns in the balance of nature and related problems in quantitative ecology.* Vol. 3. Theoretical and experimental biology. Academic Press, London and New York, 324 pp.

Received 6 June 1969

Printed 10 November 1969