



On Fitting the Poisson Lognormal Distribution to Species-Abundance Data

Author(s): M. G. Bulmer

Source: *Biometrics*, Vol. 30, No. 1 (Mar., 1974), pp. 101-110

Published by: [International Biometric Society](#)

Stable URL: <http://www.jstor.org/stable/2529621>

Accessed: 18/09/2013 10:30

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*International Biometric Society* is collaborating with JSTOR to digitize, preserve and extend access to *Biometrics*.

<http://www.jstor.org>

## ON FITTING THE POISSON LOGNORMAL DISTRIBUTION TO SPECIES-ABUNDANCE DATA

M. G. BULMER

*Department of Biomathematics, Pusey Street, Oxford, England*

### SUMMARY

An extension of MacArthur's "broken stick" model is proposed to explain why species abundances should be lognormally distributed. A method of fitting the compound Poisson lognormal distribution by maximum likelihood is described; a computer program is available for performing the calculations. It is shown how the information theory measure of species diversity can be estimated from the parameters of the fitted distribution.

### 1. INTRODUCTION

Species-abundance data are usually presented as the distribution of the number of species,  $n_r$ , represented by  $r$  members in a collection of animals or plants ( $r = 1, 2, 3, \dots$ ); the  $n_r$  are thus frequencies of frequencies. The collection is usually restricted to a particular taxonomic group (such as birds or butterflies) in a particular area. The distribution is truncated, with a missing zero class, since  $n_0$ , the number of species which are not represented at all in the collection, is usually unknown. We shall denote by  $s$  the observed number of species in the collection and by  $S$  the total number of species which might have been represented; the latter quantity is usually unknown and must be estimated from the data.

Following the treatment of Pielou ([1969] pp. 204-5), we suppose that the number of members of the  $j$ th species contained in the collection is a Poisson variate with mean  $\lambda_j$ , and that the  $\lambda_j$ 's can be regarded as constituting  $S$  independent observations from some continuous probability distribution having density function  $f(\lambda)$ . The probability that a randomly chosen species will be represented in the collection by  $r$  members is a compound Poisson distribution,

$$P_r = \int_0^\infty \frac{\lambda^r e^{-\lambda}}{r!} f(\lambda) d\lambda, \quad r = 0, 1, 2, \dots \quad (1)$$

The observed distribution can be regarded as a sample of size  $s$  from the truncated form of this distribution with the zero class missing.

In the classical approach to this problem by Fisher *et al.* [1943], it was assumed that the underlying distribution,  $f(\lambda)$ , was a Gamma distribution so that the compound distribution was negative binomial. Many observed distributions were found to be adequately fitted by the logarithmic distribution, which is the limiting form of the truncated negative binomial distribution as the shape parameter  $k$  (which is also the shape parameter of the underlying Gamma distribution) tends to zero. A stochastic model giving rise to the logarithmic distribution has been discussed by Kendall [1948]. More recently, however, it

has been suggested that it might be more appropriate to fit an underlying lognormal distribution (Preston [1948]; Williams [1964] chapter 3).

A justification of the lognormal model can be given by modifying MacArthur's "broken stick" model. MacArthur [1957] suggested that, if there are  $S$  species in a taxonomic group, their relative abundances,  $\lambda_i / \sum \lambda_i$ , might be proportional to the lengths of the segments of a line (or stick) broken at random into  $S$  pieces. This model is formally equivalent to supposing that the absolute abundances,  $\lambda_i$ , are independently and identically distributed exponential variates (Feller [1966] p. 75; Cohen [1968]), so that the compound distribution, after allowing for the Poisson sampling variability, should be negative binomial with  $k = 1$ . This is not consistent with the observation that many distributions are adequately fitted by a logarithmic distribution, which corresponds to a negative binomial with  $k \rightarrow 0$ . MacArthur himself recognized that his model could only be applied to a small, homogeneous taxonomic group, and that it broke down when applied to a larger, more heterogeneous group.

The natural way to extend the "broken stick" model to apply to a large taxonomic group is to suppose that the stick is not broken into  $S$  pieces simultaneously, but that the breakage occurs sequentially in a series of stages. Under this model we should expect a lognormal distribution of the abundances, provided that the number of stages is not too small. (See the discussion of the "theory of breakage" by Aitchison and Brown ([1957] pp. 26-7), and its application to the theory of classification and to the distribution of particle sizes resulting from the natural breakage of rocks.) It is not necessary to suppose that the breakage is random (i.e. uniform) at each stage, or even that the law of breakage is the same at different stages. The only requirement is that, at each stage, the law of breakage is independent of the size of the stick; in other words, a small stick (i.e. a rare species) is likely to be broken into the same number of pieces, of the same relative lengths, as a large stick (i.e. an abundant species). A lognormal distribution of the number of items in a class may thus be expected whenever it is natural to classify the items in a hierarchical way. This hierarchical classification may be identified either with the phylogenetic tree of the species in the group or with a classification of their ecological niches.

Previous attempts to fit a lognormal distribution to species-abundance data (Preston [1948]; Bliss [1965]) have treated the observed distribution as a truncated, grouped lognormal distribution. This treatment ignores the effect of the Poisson sampling variability. The purpose of this paper is to show how the compound Poisson lognormal distribution can be fitted by maximum likelihood. It will also be shown how the information theory measure of species diversity can be calculated from the parameters of the fitted distribution. The main computational problem is the evaluation of the probability function,  $P_r$ . A satisfactory approximation has been found for moderately large values of  $r$ , but for small values of  $r$  this function must be evaluated by numerical integration. An Algol program has been written to perform the calculations; copies are available on request.

## 2. THE POISSON LOGNORMAL DISTRIBUTION

If the abundances,  $\lambda$ , defined in the previous section are lognormally distributed, that is to say if  $\log \lambda$  is normally distributed with mean  $M$  and variance  $V$ , then the compound Poisson lognormal distribution has the probability function

$$P_r = \frac{(2\pi V)^{-1/2}}{r!} \int_0^\infty \lambda^{r-1} e^{-\lambda} e^{-(\log \lambda - M)^2 / 2V} d\lambda, \quad r = 0, 1, 2 \dots \quad (2)$$

(It should be noted that all logarithms in this paper are natural logarithms.) It is easily found by first summing over  $r$  for fixed  $\lambda$ , and then integrating that the  $k$ th factorial moment of this distribution is equal to the  $k$ th ordinary moment about the origin of the corresponding lognormal distribution; hence

$$\mu_{[k]} = \exp (kM + \tfrac{1}{2}k^2V). \tag{3}$$

The probabilities themselves are more difficult to evaluate, and there seems to be no alternative, at least for small values of  $r$ , but to determine  $P_r$  by numerical integration. It is convenient for this purpose to make the substitution  $y = \log \lambda$  so that

$$P_r = \frac{(2\pi V)^{-1/2}}{r!} \int_{-\infty}^{\infty} e^{-e^y} e^{ry} e^{-(y-M)^2/2V} dy \tag{4}$$

$$= \frac{e^{rM+r^2V/2}(2\pi V)^{-1/2}}{r!} \int_{-\infty}^{\infty} e^{-e^y} e^{-(y-M-rV)^2/2V} dy. \tag{5}$$

Some values of  $P_r$  for  $r = 0$  and 1 have been tabulated by Grundy [1951].

An approximation for large values of  $r$  can be obtained by observing that

$$P_r = \frac{(2\pi V)^{-1/2}}{r} E\{e^{-(\log X-M)^2/2V}\} \tag{6}$$

where  $X$  is a Gamma variate with parameter  $r$ , and hence with  $E(X) = V(X) = r$ . Expanding the function in curly brackets in a Taylor series about  $X = r$  and ignoring powers higher than the second, we find that

$$P_r \simeq \frac{(2\pi V)^{-1/2}}{r} e^{-(\log r-M)^2/2V} \left[ 1 + \frac{1}{2rV} \left\{ \frac{(\log r-M)^2}{V} + \log r - M - 1 \right\} \right]. \tag{7}$$

Comparison with results obtained by numerical integration shows that this approximation has a relative error less than  $10^{-3}$  when  $r \geq 10$  for values of  $M$  and  $V$  likely to be encountered in practice.

The distribution can be fitted to observed data by estimating the parameters,  $M$  and  $V$ , by the method of maximum likelihood. The calculations are described in the Appendix.

### 3. THE MEASUREMENT OF DIVERSITY

Having fitted a distribution to species abundance data, the ecologist is likely to want a measure of the ecological diversity which it represents. Many measures of diversity have been described (see Southwood [1966] chapter 13, and Pielou [1969] chapter 18, for summaries), but the most satisfactory general measure seems to be the information theory measure of diversity, defined as

$$H = - \sum_{i=1}^s p_i \log p_i \tag{8}$$

where  $p_i$  is the relative abundance of the  $i$ th species. A useful related measure is the equivalent number of species,  $S_E$ , defined as

$$S_E = e^H. \tag{9}$$

This is the number of species which would have the same value of  $H$  if they were all equally frequent. The ratio  $S_E/S$  can be used as a measure of the ‘‘equitability’’ of the observed

distribution. We shall now show how  $H$  can be estimated from the parameters of a fitted distribution.

Suppose that in the population from which the observed collection has been sampled there is a total of  $N$  individuals distributed among  $S$  species, and that there are  $N_i$  individuals in the  $i$ th species so that  $p_i = N_i/N$ . The amount of information is

$$H = - \sum_{i=1}^S \frac{N_i}{N} \log p_i . \quad (10)$$

If we take each individual in turn and record the value of  $p_i$  for the species to which it belongs, then

$$H = - \frac{1}{N} \sum_{k=1}^N \log p_k \quad (11)$$

the summation now being taken over individuals. Hence

$$H = -E(\log p) \quad (12)$$

where the distribution of  $p$  is that obtained in sampling individuals at random. If the underlying distribution of  $p$  among the species has density function  $f(p)$ , then the density function of  $p$  in sampling individuals at random is given by

$$g(p) = pf(p) / \int_0^1 pf(p) dp . \quad (13)$$

(This formula expresses the fact that common species contain more individuals than rare species; sampling by individuals is equivalent to sampling by species with probability proportional to their abundance,  $p$ . The denominator is a normalizing factor.) This distribution is the first moment distribution of  $f(p)$  as defined by Aitchison and Brown ([1957] p. 12).

To apply this result, let us first suppose that the  $\lambda_i$ 's are sampled from a Gamma distribution with parameter  $k$ . Since

$$p_i = \lambda_i / \sum_{i=1}^S \lambda_i \quad (14)$$

the  $p_i$ 's can be regarded as observations from a Beta distribution with parameters  $k$  and  $k(S-1)$  subject to the constraint that  $\sum p_i = 1$ . The distribution  $g(p)$  is thus a Beta distribution with parameters  $k+1$  and  $k(S-1)$ . Hence

$$H = \psi(kS+1) - \psi(k+1) \quad (15)$$

where  $\psi(\cdot)$  is the digamma function. (See Kendall and Stuart [1963] p. 36 for the expectation of the logarithm of a Beta variate.) As  $k \rightarrow 0$ , it is quite easy to show that  $S$  must tend to infinity in such a way that  $kS = \alpha$ , the constant in the logarithmic distribution often used as a measure of species diversity. (This follows from the fact that  $k\Gamma(k) \rightarrow 1$  as  $k \rightarrow 0$ .) For data fitted by the logarithmic distribution we may thus estimate  $H$  as

$$\hat{H} = \psi(\hat{\alpha}+1) - \psi(1) = \psi(\hat{\alpha}+1) + \gamma \quad (16)$$

where  $\gamma = 0.57722 \dots$  is Euler's constant. If  $\hat{\alpha}$  is not too small, we can to a good approximation write  $\psi(\hat{\alpha}+1) = \log(\hat{\alpha} + \frac{1}{2})$ . (Numerical investigation has shown this to be considerably better than the usual approximation,  $\psi(x) = \log x$ .) The sampling variance of

this estimated value of  $H$  is approximately

$$\text{var}(\hat{H}) = \psi'^2(\alpha + 1) \text{var}(\hat{\alpha}) \simeq \text{var}(\hat{\alpha})/(\alpha + \tfrac{1}{2})^2 \quad (17)$$

where  $\psi'(\cdot)$  is the trigamma function. The equivalent number of species is

$$\hat{S}_E = e^{\hat{H}} = 1.781e^{\psi'(\hat{\alpha}+1)} \simeq 0.89 + 1.781\hat{\alpha}. \quad (18)$$

The standard error of this quantity is approximately the standard error of  $\hat{H}$  multiplied by  $\hat{S}_E$ .

If the  $\lambda_i$ 's are lognormally distributed, it is not easy to determine the exact distribution of the  $p_i$ 's from equation 14, but a reasonable approximation should be obtained by assuming that  $\sum \lambda_i$  is equal to its expected value,  $S \exp(M + \frac{1}{2}V)$ , so that  $p$  is lognormally distributed with parameters  $(-\log S - \frac{1}{2}V)$  and  $V$ . Hence  $g(p)$  is lognormal with parameters  $(-\log S + \frac{1}{2}V)$  and  $V$ , since the first moment distribution of a lognormal distribution with parameters  $M$  and  $V$  is also lognormal with parameters  $(M + V)$  and  $V$  (Aitchison and Brown [1957] p. 12). Thus

$$\hat{H} = \log \hat{S} - \tfrac{1}{2}\hat{V} \quad (19)$$

where  $\hat{S}$  is estimated from equation A-11 of the Appendix. After taking into account the correlation between  $\hat{P}_0$  and  $\hat{V}$ , the sampling variance of  $\hat{H}$  is approximately given by

$$\begin{aligned} \text{var}(\hat{H}) &= \frac{\text{var}(\hat{S})}{\hat{S}^2} + \tfrac{1}{4} \text{var}(\hat{V}) \\ &+ \frac{1}{(1 - P_0)} \{P_1 \text{cov}(\hat{H}, \hat{V}) + \tfrac{1}{2}(P_1 - 2P_2) \text{var}(\hat{V})\}. \end{aligned} \quad (20)$$

For a nontruncated distribution  $S$  is known so that

$$\text{var}(\hat{H}) = \tfrac{1}{4} \text{var}(\hat{V}). \quad (21)$$

# EXAMPLES

As an example we shall consider first the well-known data of Williams on the numbers of Macrolepidoptera caught in a light trap at Rothamsted, England. These data have been analysed by Bliss [1965] who fitted both a logarithmic distribution and a truncated, grouped lognormal distribution to them. His results for the year 1934 are reproduced in Table 1, together with the Poisson lognormal distribution fitted by maximum likelihood. It will be seen that both the logarithmic and the lognormal models provide a satisfactory fit, and that there is little difference between the grouped lognormal distribution fitted by Bliss and the Poisson lognormal distribution fitted by maximum likelihood. The estimates of the parameters of the fitted distributions and other figures derived from them are shown in Table 2. The estimates of  $M$  and  $V$  calculated by Bliss differ somewhat from those calculated by maximum likelihood; this has affected the estimated total number of species but has made little difference to the estimate of species diversity. The estimate of species diversity calculated from the logarithmic distribution is considerably higher.

For another example we consider the data of Corbet [1942] on Malayan butterflies, which have also been considered by Fisher *et al.* [1943] and by Williams ([1964] pp. 19–21). These data are censored, since when 25 individuals had been caught little effort was made to obtain others of the same species. Corbet caught altogether 620 species, but he also estimated that the total butterfly fauna of the area contained 924 species, so that 304

TABLE 1  
DISTRIBUTION OF LEPIDOPTERA CAUGHT IN A LIGHT TRAP AT ROTHAMSTED IN 1934

Individuals per species (r)	Number of species (n <sub>r</sub> )	Expected		
		Logarithmic	Grouped lognormal	Poisson lognormal
1	34	39.0	32.7	31.2
2	19	19.3	20.5	20.8
3	15	12.7	14.6	15.0
4	10	9.4	11.2	11.5
5	10	7.4	8.9	9.1
6-7	9	11.3	13.4	13.7
8-10	17	12.0	13.7	13.9
11-14	9	11.0	12.0	12.1
15-20	14	11.2	11.3	11.3
21-28	10	9.8	9.1	9.0
29-39	6	8.7	7.4	7.3
40-55	7	7.8	6.1	6.0
56-77	3	6.2	4.6	4.5
78-108	5	4.6	3.4	3.4
109-151	4	2.8	2.4	2.4
152-	4	2.9	4.5	4.7
Total	176	176.1	175.8	175.9
χ <sup>2</sup>		9.2	6.9	7.4
Degrees of freedom		14	13	13

species were missing from the collection. The data are shown in Table 3, together with the Poisson lognormal distribution fitted by maximum likelihood both omitting and including the zero class. The estimated parameters and other derived figures are shown in Table 4. A more satisfactory fit is obtained with the zero class missing, and the estimate

TABLE 2  
PARAMETER ESTIMATES FOR THE DATA OF TABLE 1 WITH THEIR STANDARD ERRORS

	Logarithmic	Grouped lognormal	Poisson lognormal
$\hat{\alpha}$	39.4 ± 1.5		
$\hat{M}$		1.45 ± 0.21	1.17 ± 0.28
$\hat{V}$		3.18 ± 0.54	3.56 ± 0.68
Cov( $\hat{M}, \hat{V}$ )			-0.140
Total number of species, $\hat{S}$	∞	198.8	226.4 ± 16.8
$\hat{H}$	4.26 ± 0.04	3.70	3.64 ± 0.29
Equivalent number of species, $\hat{S}_E$	71.1 ± 2.7	40.4	38.1 ± 11.1

TABLE 3  
DISTRIBUTION OF CORBET'S MALAYAN BUTTERFLIES

Individuals per species (r)	Number of species (n <sub>r</sub> )	Expected (Poisson lognormal)	
		Omitting zero class	Including zero class
0	(304)	(195.3)	295.0
1	118	110.2	127.4
2	74	71.3	74.6
3	44	50.9	50.7
4	24	38.7	37.5
5	29	30.7	29.3
6	22	25.2	23.7
7	20	21.1	19.7
8	19	18.0	16.7
9	20	15.6	14.4
10	15	13.6	12.6
11	12	12.1	11.1
12	14	10.8	9.9
13	6	9.7	8.9
14	12	8.8	8.1
15	6	8.0	7.3
16	9	7.3	6.7
17	9	6.7	6.2
18	6	6.2	5.7
19	10	5.7	5.3
20	10	5.3	4.9
21	11	4.9	4.6
22	5	4.6	4.3
23	3	4.3	4.0
24	3	4.1	3.8
25+	119	126.2	131.3
Total	620 (924)	620.0	923.7
$\chi^2$		30.5	36.8
Degrees of freedom		22	23
Probability		0.107	0.034

TABLE 4  
PARAMETER ESTIMATES FOR THE DATA OF TABLE 3 WITH THEIR STANDARD ERRORS

	Omitting zero class	Including zero class
$\hat{M}$	1.13 ± 0.18	0.70 ± 0.09
$\hat{V}$	4.11 ± 0.59	5.42 ± 0.43
Cov ( $\hat{M}, \hat{V}$ )	-0.086	-0.011
Total number of species, $\hat{S}$	815.30 ± 40.5	924
$\hat{H}$	4.65 ± 0.26	4.12 ± 0.22
Equivalent number of species, $\hat{S}_E$	104.50 ± 26.8	61.60 ± 13.3



of the total number of species is rather lower than predicted, but is nevertheless of the right order of magnitude.

# SUR L'AJUSTEMENT D'UNE DISTRIBUTION LOGNORMALE DE POISSON A DES DONNEES D'ABONDANCE DES ESPECES.

## RESUME

On propose une extension du modèle "broken stick" de MacArthur pour expliquer pourquoi des abondances d'espèces doivent être de distribution lognormale. On décrit une méthode d'ajustement à une distribution lognormale de Poisson composée par le maximum de vraisemblance; un programme est disponible pour effectuer les calculs.

Il est montré comment la mesure de théorie de l'information de la diversité des espèces peut être estimée à partir des paramètres de la distribution ajustée.

## REFERENCES

- Aitchison, J. and Brown, J. A. C. [1957]. *The Lognormal Distribution*. Cambridge University Press.
- Bliss, C. I. [1965]. An analysis of some insect trap records. In: *Classical and Contagious Discrete Distributions*. Patil, G. P. (Ed.), pp. 385-97. Pergamon, Oxford.
- Cohen, J. E. [1968]. Alternate derivations of a species-abundance relation. *Amer. Nat.* 102, 165-72.
- Corbet, A. S. [1942]. The distribution of butterflies in the Malay Peninsula. *Proc. R. Ent. Soc. Lond. (A)* 16, 101-16.
- Feller, W. [1966]. *An Introduction to Probability Theory and its Applications, Vol. 2*, Wiley, New York.
- Fisher, R. A., Corbet, 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.
- Grundy, R. M. [1951]. The expected frequencies in a sample of an animal population in which the abundances of species are log-normally distributed. *Biometrika* 38, 427-34.
- Kendall, D. G. [1948]. On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika* 35, 6-15.
- Kendall, M. G. and Stuart, A. [1963]. *The Advanced Theory of Statistics, Vol. 1*, (2nd Ed.), Griffin, London.
- MacArthur, R. H. [1957]. On the relative abundance of bird species. *Proc. Nat. Acad. Sci. Wash.* 43, 293-5.
- Pielou, E. C. [1969]. *An Introduction to Mathematical Ecology*. Wiley, New York.
- Preston, F. W. [1948]. The commonness and rarity of species. *Ecology* 29, 254-83.
- Southwood, T. R. E. [1966]. *Ecological Methods*. Methuen, London.
- Williams, C. B. [1964]. *Patterns in the Balance of Nature*. Academic Press, London and New York.

## APPENDIX

### MAXIMUM LIKELIHOOD ESTIMATION

If we consider the observed data as a sample of size  $s$  from a truncated Poisson log-normal distribution, the log likelihood is

$$L = \sum_{r=1} n_r \log \{P_r / (1 - P_0)\} = \sum_{r=1} n_r \log P_r - s \log (1 - P_0). \quad (A1)$$

If we write  $\theta_1 = M$ ,  $\theta_2 = V$ , the ML score with respect to  $\theta_i$  is

$$\frac{\partial L}{\partial \theta_i} = \sum_{r=1} \frac{n_r}{P_r} \frac{\partial P_r}{\partial \theta_i} + \frac{s}{(1 - P_0)} \frac{\partial P_0}{\partial \theta_i}. \quad (A2)$$

Differentiating again and taking expected values, we find that the  $(i, j)$ th element of the information matrix is

$$-E\left(\frac{\partial^2 L}{\partial \theta_i \partial \theta_i}\right) = \frac{s}{(1 - P_0)} \left\{ \sum_{r=1} \frac{1}{P_r} \frac{\partial P_r}{\partial \theta_i} \frac{\partial P_r}{\partial \theta_i} - (1 - P_0)^{-1} \frac{\partial P_0}{\partial \theta_i} \frac{\partial P_0}{\partial \theta_i} \right\}. \quad (\text{A3})$$

To evaluate  $\partial P_r / \partial \theta_i$ , we use the expression for  $P_r$  given in equation 4. Integration by parts shows that, for all values of  $r$ ,

$$\frac{\partial P_r}{\partial M} = rP_r - (r + 1)P_{r+1} \quad (\text{A4})$$

$$\frac{\partial P_r}{\partial V} = \frac{1}{2} \frac{\partial^2 P_r}{\partial M^2} = \frac{1}{2} \{ r^2 P_r - (r + 1)(2r + 1)P_{r+1} + (r + 1)(r + 2)P_{r+2} \}.$$

The ML scores (equation A2) can now be evaluated. The infinite sum in equation A3 can be evaluated directly for  $r = 1$  to  $R$ , say. The remainder, for  $r > R$ , can be evaluated approximately by assuming that

$$P_r \simeq \frac{(2\pi V)^{-1/2}}{r} e^{-(\log r - M)^2 / 2V} \quad (\text{A5})$$

so that  $\partial \log P_r / \partial \theta_i$  is a linear or quadratic function of  $(\log r - M)$ . We next observe that

$$\sum_{r=R+1}^{\infty} \frac{1}{P_r} \frac{\partial P_r}{\partial \theta_i} \frac{\partial P_r}{\partial \theta_i} = \sum_{r=R+1}^{\infty} P_r \frac{\partial \log P_r}{\partial \theta_i} \frac{\partial \log P_r}{\partial \theta_i}. \quad (\text{A6})$$

By approximating the sum from  $R + 1$  to infinity by the corresponding integral from  $R + \frac{1}{2}$  to infinity and then making the substitution  $y = (\log r - M)$ , the quantity (A6) can be expressed to a good approximation in terms of the moments of a truncated normal distribution.

The above argument can be modified in an obvious way if the zero class can be observed so that the distribution is not truncated. It may also happen that the distribution is censored to the right, so that the frequencies  $n_1, n_2, \dots, n_R$  are known for  $r \leq R$ , but only the total number of species with more than  $R$  members is known. If we denote this number by  $n^*$  and if we denote the corresponding (nontruncated) probability by  $P^*$ , then the log likelihood is

$$L = \sum_{r=1}^R n_r \log P_r - s \log (1 - P_0) + n^* \log P^* \quad (\text{A7})$$

in the truncated case; in the nontruncated case the summation in the first term extends from  $r = 0$  to  $R$  and the second term is missing. The ML scores must be modified by terminating the summation at  $r = R$  and adding the term

$$\frac{n^*}{P^*} \frac{\partial P^*}{\partial \theta_i}. \quad (\text{A8})$$

The information matrix is modified by substituting

$$\frac{1}{P^*} \frac{\partial P^*}{\partial \theta_i} \frac{\partial P^*}{\partial \theta_i} \quad (\text{A9})$$

for the remainder given by equation A6. It is easily verified from equation A4 that

$$\frac{\partial P^*}{\partial M} = (R + 1)P_{R+1} \quad (\text{A10})$$

$$\frac{\partial P^*}{\partial V} = \frac{1}{2}(R + 1)[(R + 1)P_{R+1} - (R + 2)P_{R+2}].$$

In the case of a truncated distribution the total number of species (including the missing ones) can be estimated as

$$\hat{S} = s/(1 - \hat{P}_0). \quad (\text{A11})$$

The sampling variance of this estimate is approximately

$$\begin{aligned} \text{var}(\hat{S}) &= \frac{\text{var}(s)}{(1 - P_0)^2} + s^2 \text{var}\left(\frac{1}{1 - \hat{P}_0}\right) \\ &= \frac{\text{var}(s)}{(1 - P_0)^2} + \frac{s^2}{(1 - P_0)^4} \text{var}(\hat{P}_0). \end{aligned} \quad (\text{A12})$$

Furthermore,

$$\begin{aligned} \text{var}(s) &= SP_0(1 - P_0) \simeq sP_0 \\ \text{var}(\hat{P}_0) &= \left(\frac{\partial P_0}{\partial M}\right)^2 \text{var}(\hat{M}) + 2 \frac{\partial P_0}{\partial M} \frac{\partial P_0}{\partial V} \text{cov}(\hat{M}, \hat{V}) + \left(\frac{\partial P_0}{\partial V}\right)^2 \text{var}(\hat{V}) \\ &= P_1^2 \text{var}(\hat{M}) + P_1(P_1 - 2P_2) \text{cov}(\hat{M}, \hat{V}) + \frac{1}{4}(P_1 - 2P_2)^2 \text{var}(\hat{V}). \end{aligned} \quad (\text{A13})$$

*Received November 1972, Revised June 1973*

*Key Words:* Species abundance; Poisson lognormal distribution; Broken stick model.