

402

**STATISTICAL MECHANICAL APPROACH  
TO MODELS OF ECOSYSTEMS**

Dissertation submitted to the Jawaharlal Nehru University  
in partial fulfilment of the requirements for the award  
of the Degree of  
**MASTER OF PHILOSOPHY**

74/12

**N. GUNASEKARAN**

SCHOOL OF ENVIRONMENTAL SCIENCES  
JAWAHARLAL NEHRU UNIVERSITY  
NEW DELHI-110067, INDIA  
**1981**

C E R T I F I C A T E

The work presented in this Thesis has been carried out in the School of Environmental Sciences, Jawaharlal Nehru University, New Delhi. The work is original and has not been submitted so far in part or full for any other degree or diploma of any University.

*N. Gunasekaran*  
N. GUNASEKARAN  
(Candidate)

*L. K. Pande*  
L. K. PANDE  
(Supervisor)

*B. Shetty*  
(Dean) Dean  
School of Environmental Sciences  
Jawaharlal Nehru University  
New Delhi-110057

Date: 10-12-1981

SCHOOL OF ENVIRONMENTAL SCIENCES  
JAWAHARLAL NEHRU UNIVERSITY  
NEW DELHI - 110 067.

## A C K N O W L E D G E M E N T

The work contained in the present Thesis was suggested and guided by my Supervisor Dr. L.K. Pande. It is a pleasure to record my deep sense of gratitude to him for his persistent guidance and delightful discussions. I also acknowledge having benefited immensely from his lectures on the subject of 'Mathematical Ecology' and 'Statistical Mechanics'.

I take this opportunity to express my heartfelt thanks to Dr. J. Subba Rao and Dr (Mrs) Nalini Bhat, from whom I benefited in discussions.

I wish to thank many of my friends and colleagues, especially C.P. Geevan, J.A. Maliekal, M. Krishna and C.D. Thakur with whom I had useful discussions. Thanks are due to S.G. Krishnan, D.S. Kumar and Aruna, who helped me one way or other, to complete this work.

Thanks are also due to Gajanan for typing the manuscript.

I acknowledge the award of Junior Research Fellowship by the Jawaharlal Nehru University.

December 1981

NEW DELHI - 110 067.

*N. Gunasekaran*  
[N. GUNASEKARAN]

## C O N T E N T S

	<u>Page No.</u>
CHAPTER - I : INTRODUCTION	1
CHAPTER - II : ON RELATIVE ABUNDANCE OF SPECIES	10
CHAPTER - III: PRESTON'S ANALYSIS	25
CHAPTER - IV : STATISTICAL MECHANICAL APPROACH TO THE LOTKA- VOLTERRA MODEL	39
CHAPTER - V : STATISTICAL MECHANICAL APPROACH TO THE GOMPERTZ MODEL	60
REFERENCES	74

## CHAPTER - I

### INTRODUCTION

Most of the genera in ecological systems contain many species of organisms and these species vary greatly in their abundance from very common to very rare. Frequently, it is found through experimental collections that singleton species are numerous, and species with successively more representatives, doubletons with two, trebletons with three, ....., and so on, are usually progressively less numerous. This relative abundance of different species within a genera is repeatedly observed by taxonomists experimentally. Some of the well known observations, for example, are:

1. Corbet's results<sup>(1)</sup> on Malayan butterfly flies;
2. Williams' results<sup>(2)</sup> on moth species obtained by means of a light trap at Rothamsted;
3. Saunders' observations<sup>(3)</sup> on the birds of Quaker Run Valley, western New York State;

4. Preston and Norris' data<sup>(4)</sup> on the Breeding birds of the Frith (Preston Laboratory grounds);
5. Dirk's data<sup>(5)</sup> on Moths in a light trap at Orono, Maine, and
6. Seamans' data<sup>(6)</sup> on moths in a light trap at Lethbridge, Alberta.

As soon as one attempts to study these data, the question immediately arises: how are the abundances of the different species distributed? Attempts to answer this question have led to the development of few species abundance relations and other related material. A brief summary is as follows:

I. Corbet<sup>(1)</sup> gave the relation between the number of species  $S$  and the number of individuals  $n$  in each of the  $S$  species as

$$S = \frac{C}{n^m} \dots \dots \dots [1]$$

where,  $C$  and  $m$  are constants.

II. Williams<sup>(2)</sup> suggested the use of the series

$$N \left( 1 + \frac{1}{2} + \frac{1}{3} + \dots \dots \dots \right) \dots \dots \dots [2]$$

where the terms give the number of species represented by 1, 2, 3, ....., individuals (or the

frequencies of different species), respectively.

III. Fisher<sup>(7)</sup> gave the distribution

$$S = -\alpha \log(1-x) = \alpha x + \frac{\alpha x^2}{2} + \frac{\alpha x^3}{3} + \dots [3a]$$

and

$$N = \frac{\alpha x}{1-x} = \alpha x + \alpha x^2 + \alpha x^3 + \dots [3b]$$

The different terms in the series [3a] give the frequencies of different species. S is here the total number of species and N is the total number of individuals in the collection, and  $\alpha$  and  $x$  are constants characteristic of the collection. Once S and N are known for a particular collection,  $\alpha$  and  $x$  can be calculated using equations [3] and the series giving the frequencies of different species is then known for that collection. Thus, the salient feature of Fisher's distribution is that for any collection whose total number of species S and the total number of individuals N are known, the series giving the frequencies of different species can be predicted.

IV<sub>b</sub> Preston<sup>(8)</sup> proposed the distribution

$$n = n_0 e^{-(aR)^2} \quad \dots \dots \dots [4]$$

where  $n_0$  is the number of species belonging to the modal octave,  $n$  is the number of species belonging to an octave distance,  $R$  octaves from the modal octave and 'a' is a constant calculated from the experimental data. As noted in Chapter III, the terms "octave" and "model octave" have a very simple meaning. It will also be seen from the discussion in that Chapter that Preston graduated the abscissa in his plots as the logarithm of the number of individuals per species and the ordinate as the frequencies of different species. Therefore, the expression [4] given by Preston<sup>(8)</sup> is actually a log normal distribution.

Now, as regards the suggestion of Corbet<sup>(1)</sup>, we note that Corbet himself pointed out that his



relation cannot be a true representative for any random sample, because his sample was biased in the direction of too many species with small numbers, and too small numbers of individuals for the commoner species.

The difficulty with the suggestion of Williams<sup>(2)</sup> is that the sum of all the terms in his series is infinite, calling for an infinite number of species in the collection. Fisher recognized this and gave a distribution based on a more sophisticated mathematical approach. He started with the poisson distribution by taking the number of individuals of any species as the poisson variate. Since the collection contained a heterogeneous population, the mean of the poisson was not a true mean of the population and that itself called for a distribution. Therefore, Fisher<sup>(7)</sup> assumed  $\chi^2$  distribution for the mean of the poisson. The resulting distribution ~~for~~ thus became a compound poisson distribution. This resulted in the series set out in equation [3], which fitted the observed data.

Now, the Fisher series explained the Williams data satisfactorily and also essentially

Corbet's<sup>(1)</sup> data, barring few deviations. But the detailed investigations of Preston<sup>(8)</sup>, which were carried out subsequently, lead to a different picture. Preston analysed six different sets of data, each from a different genera, and observed that all of them followed a log normal curve. Therefore, it is implied from this that the distribution which is actually the more extensively valid one, is not the Fisher distribution, but rather the standard log normal distribution given by Preston.

It is of much interest to develop a conceptual basis, either in the form of a complete theory or at least as a theoretical model, from which the distribution followed by the relative abundance of species within a genera could be properly understood. If this could be done much light would be thrown into the nature of multi-species systems and their interactions, which in turn would enable us to understand ecosystems and their course of evolution on a precise quantitative basis.

A first attempt in this direction was made by Kendall<sup>(9)</sup>. He considered a simple stochastic model which allowed for the birth and death as well as the immigration processes for the population. He was able to obtain a probability distribution for the population which was analogous to that of Fisher given in equations [3] above. This result was, however, obtained for a single species model.

The problem was taken up within a multi-species framework by Kerner<sup>(10)</sup>. He assumed the dynamics of the multispecies system to be given by the coupled Lotka-Volterra equations. Under the assumption of a constraint, he was able to develop the statistical mechanics of these equations. This then led to a population distribution for each of the component species which was precisely the  $\chi^2$ -distribution assumed by Fisher for the so-called intrinsic abundance of species which he fed into his poisson distribution.

Significant as it is, the work of Kerner raises two problems. One is that it does not lead to a full explanation of the Fisher result

as given by equations [3] above, but leads only to an understanding of an intermediate result. The second problem is that, in view of the results of Preston, the distribution which needs to be explained is not the  $\chi^2$  or the compound poisson distribution of Fisher, but the lognormal distribution given by equation [4].

In the present thesis, we consider this problem afresh. Our starting point is to assume that the multispecies ecosystems which are of interest to us may be described by a set of equations which contain interactions between species in the Gompertz form. Models characterised by such interactions have been discussed for two and three species systems earlier<sup>(11)</sup>. We next develop the statistical mechanics of our equations in a manner which is analogous to that of Kernar. It will be seen that the population distributions that we obtained from here are precisely of the lognormal form. For details and concluding remarks on this matter, we refer to the last Chapter of this thesis.

The material presented in this thesis is planned as follows. In Chapter II, we discuss the experimental and theoretical analysis arising out of the works of Corbet, Williams and Fisher. Chapter III is devoted to the results of Preston. In Chapter IV we develop, following Kerner, the statistical mechanics of the Lotka-Volterra model leading to the  $\chi^2$  distribution required by Fisher. Finally, in Chapter V, we discuss our own multi-species model and develop its statistical mechanics, which leads to the lognormal distribution suggested by the analysis of Preston.

CHAPTER - IION RELATIVE ABUNDANCE OF SPECIES

Any animal or plant population in the wild state consists basically of a very large number of individuals. The individuals are classified by naturalists into species. The number of individuals within a species is changing with time because of the intra- and inter-species interactions. As a result of these interactions, each species establishes an uneasy balance of numbers among all the others, resulting in a pattern of relative abundance for different species in the given population.

It has been found that if a random sample of individuals is taken from a mixed wild population of animals containing a large number of species, there appears to be a mathematical order in the relative abundance of the different species represented. In general more species are represented by one individual than by two, more by two than by three, and so on. Since order cannot be maintained out of chaos by the mere process of sampling, we must infer that there is some order in the relative abundance of the species in the population itself.

The first object of the present study, discussed in this and the next Chapter, is to review the existing literature which establishes the above mentioned order in the relative abundance of different species in any population. While doing this review, we also discuss certain mathematical series that have been suggested at various times to fit the observed data and to see what light the structure of these series can throw on the pattern of frequency distributions in the populations sampled.

#### 1. Corbet's Observations

In 1942, Corbet<sup>(1)</sup> pointed out that in a large collection of butterflies which he had made in Malaya, the 9031 individuals included 620 species. Of these, 118 species were each represented by only a single individual (called singletons), seventy-four were represented by two individuals (doubletons), forty-four by three individuals, and so on.

An attempt was then made by Corbet to find a mathematical expression that would fit this data. He found a relation between  $S$  which is the number of species and  $n$  which is the number of individuals per species in any of the  $S$  species. The relation is,

$$S = \frac{C}{n^m} \dots \dots \dots [1]$$

where m and c are constants. Equation [1] can also be written as

$$\log^S + m \log^N = \log^C \dots \dots \dots [2]$$

This latter is a straight line equation.

Corbet himself pointed out that his sample was not a truly random one, as more attention was paid to the catching of a rare species than a common one. When about twenty-five individuals of any one species had been captured, little effort was made to obtain others. Thus, the sample was biased in the direction of too many species with small numbers, and too small numbers of individuals for the commoner species. Hence equation [1] cannot be a true distribution for a random sample.

## 2. Williams' Observations

C.B. Williams<sup>(2)</sup> conducted random sampling of Lepidoptera by means of a light-trap at Rothamsted Experimental Station, about 25 miles north of London. His experimental data were published in the group of papers by Corbet, Williams and Fisher (1943).<sup>(7)</sup> He was struck by the fact that there were usually great numbers of species represented by singletons, a much less number of species represented by doubletons and



still fewer species represented by three, four and so on. He also noticed that if  $N$  be the number of species represented by singletons, the number represented by doubletons was  $\frac{N}{2}$ , the number represented by three was  $\frac{N}{3}$  and so on. Hence in the collection, if the observations were valid indefinitely, we should have a series

$$N \left( 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots \right) \dots [3]$$

where the different terms represent, respectively, the number of species represented by

1, 2, 3, 4, . . . . ., individuals.

This series does not conform well to the experimental observations when the size of the sample is increased. Furthermore, this being a harmonic series, the sum of its terms is infinite, calling for an infinite number of species in the collection. Williams himself recognized this. It is clear from here that the interpretation of the frequency distribution of species and individuals requires a more sophisticated theoretical and mathematical reasoning.

### 3. The Use of the Poisson Distribution

So, we begin at a more basic level. For this, consider any random sample from a homogeneous ecosystem. Let each of these samples have the capacity to contain

$N$  individuals in total. Now the probability of observing  $n$  individuals of a species in any sample is given by the binomial distribution

$$P(n) = {}^N C_n (P)^n (1-P)^{N-n} \dots \dots \dots [4]$$

where  $P$  is the probability of finding a single individual in the sample. Now if  $N$  is very large and  $P$  is very small, but that the mean of this distribution  $NP \equiv m$  is of moderate magnitude, and furthermore, if  $n$  is negligibly small compared to  $N$ , the expression

$$\begin{aligned} P(n) &= {}^N C_n (P)^n (1-P)^{N-n} \\ &= N(N-1)(N-2)\dots(N-n+1) \left(\frac{m}{N}\right)^n \frac{(1-\frac{m}{N})^{N-n}}{n!}, \end{aligned}$$

which may be written in the form

$$\begin{aligned} P(n) &= \frac{m^n}{n!} \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \dots \\ &\dots \left\{1 - \frac{(n-1)}{N}\right\} \left\{1 - \frac{m}{N}\right\}^{N-n} [5] \end{aligned}$$

goes over, as  $N$  becomes indefinitely large, to

$$P(n) \longrightarrow \frac{m^n e^{-m}}{n!} \dots \dots \dots [6]$$

This is the well known Poisson distribution with  $n$ , the number observed in any sample, as the poisson variate and  $m$  as the mean or the expectation value of the distribution.

4. The Analysis of Fisher

The above distribution was the starting point for Fisher's analysis for the interpretation of Corbet's and Williams' data<sup>(1,2)</sup>. According to him, if successive, independent, equal samples be taken from a homogeneous ecosystem, the numbers observed can be related to the poisson series

$$\frac{e^{-m} m^n}{n!} \dots \dots \dots [7]$$

where, as noted earlier, m is the expected number and n is the number observed in any sample. In general the probability that a species will be represented by n individuals is then

$$P_n = \frac{e^{-m} m^n}{n!} \dots \dots \dots [8]$$

$$n = 0, 1, 2, \dots \dots \dots$$

Now, in the present distribution, m is the mean number of individuals per species. But in the actual population, the number of individuals per species varies from a very small number, say, of the order of one to a very large number, of the order of lakhs and millions. Hence the quantity m which is the expectation value of the Poisson distribution cannot be the expectation for all the species. That is, if the material sampled were heterogeneous or if unequal samples were taken, we should have a mixture of distributions

corresponding to different values of  $m$ . The same is true of the numbers of different organisms observed in a single sample, if the different species are not equally abundant. Hence if equation [8] is to explain the species - individuals pattern correctly, the parameter  $m$  itself may follow a distribution.

Therefore, for the variable  $m$ , Fisher <sup>(7)</sup> assumed the well known  $\chi^2$  distribution, which is of the form

$$df = \frac{1}{\Gamma(k)} p^{-k} m^{k-1} e^{-m/P} dm, \dots \dots [9]$$

$df$  here being the element of frequency or probability with which  $m$  falls in the infinitesimal range  $dm$ . The parameters  $P$  and  $k$  are related by  $\frac{M}{K} = P$ , where  $M$  is the mean value of  $m$  and  $k$  measures inversely the variability of  $m$ . Thus a large value of  $k$  would be expected if the densities of various species differed only slightly from one another, and a small value of  $k$  would be expected if there were pronounced differences in their densities.

If we multiply this expression [9] by the expression given in [8], which is the probability of observing just  $n$  organisms, and integrate with respect to  $m$  over its whole range from 0 to  $\infty$ , we will have

$$\begin{aligned}
P_n &= \int_0^{\infty} \frac{1}{(k-1)!} p^{-k} e^{-m/p} \frac{e^{-m} m^{n+k-1}}{n!} dm \\
&= \frac{p^{-k}}{(k-1)! n!} \int_0^{\infty} m^{k+n-1} e^{-m(1+\frac{1}{p})} dm \\
&= \frac{p^{-k}}{(k-1)! n!} \int_0^{\infty} m^{k+n-1} e^{-m \left(\frac{p+1}{p}\right)} dm \dots [10]
\end{aligned}$$

Putting  $m \left(\frac{p+1}{p}\right) = \phi$ ,

$$dm = \frac{d\phi}{\left(\frac{p+1}{p}\right)},$$

and  $m^{k+n-1} = \frac{\phi^{k+n-1}}{\left(\frac{p+1}{p}\right)^{k+n-1}}$ ,

and substituting these into equation [10], we have

$$\begin{aligned}
P_n &= \frac{p^{-k}}{(k-1)! n! \left(\frac{p+1}{p}\right)^{k+n}} \int_0^{\infty} e^{-\phi} \phi^{k+n-1} d\phi \\
&= \frac{p^{-k} \Gamma(k+n)}{(k-1)! n! \left(\frac{p+1}{p}\right)^{k+n}},
\end{aligned}$$

so that we finally get

$$P_n = \frac{(k+n-1)! p^n}{(k-1)! n! (p+1)^{k+n}} ; \dots [11]$$

where  $n$  is now a negative binomial variate. This nomenclature arises from the similarity which the above expression bears with the negative binomial expansion

$$\left\{ 1 - \frac{P}{1+P} \right\}^{-k} = \sum_{n=0}^{\infty} \frac{(k+n-1)!}{(k-1)! n!} \left\{ \frac{P}{1+P} \right\}^n .$$

For convenience, let us introduce the variable

$$X = \left( \frac{P}{1+P} \right) \text{ such that } 0 < X < 1.$$

Then,

$$P_n = \frac{(k+n-1)! (1-X)^k X^n}{(k-1)! n!}$$

Since we are not interested in the number of species represented by zero members in the collection, the zero class denoted by  $P_0$  is to be excluded, where

$$P_0 = (1 - X)^k \quad . \quad . \quad . \quad . \quad . \quad [12]$$

We note that

$$\sum_{n=0}^{\infty} P_n = 1.$$

If we exclude  $P_0$  from the sum here, the right hand side is no more unity. But we do have

$$\frac{(P_1 + P_2 + \dots \dots \dots)}{(1 - P_0)} = 1.$$

The exclusion of  $P_0$  from the sum thus necessitates the use of

$$P_n^1 = \frac{P_n}{1 - P_0}, \quad n = 1, 2, \dots \quad [13]$$

which are now the properly normalised quantities. By collecting terms independent of  $n$  into a single constant  $C$ , we can write

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

where,

$$C = \frac{(1-x)^k}{[1 - (1-x)^k]} \frac{1}{\Gamma(k)}.$$

In natural communities it is usually found that the differences in abundance among the species are extremely great. This led Fisher, to propose that by letting  $k \rightarrow 0$  in the formula for  $P_n^1$  an approximation to species-abundance proportions might be obtained.

Allowing  $k \rightarrow 0$  in equation [13] to give the limiting form of  $P_n^1$ , we get

$$\pi_n \equiv \lim_{k \rightarrow 0} P_n^1 = \gamma \frac{\Gamma(n)}{n!} x^n = \gamma \frac{x^n}{n}$$

for  $n = 1, 2, 3, \dots$

where,  $\gamma = \lim_{k \rightarrow 0} C$

$\gamma$  can be evaluated by noting that

$$\sum_{n=1}^{\infty} \pi_n = -\gamma \log(1-x) = 1$$

$$\therefore \gamma = \frac{-1}{\log(1-x)}$$

The expression  $\frac{\gamma x^n}{n}$  is the probability that a species will be represented in the collection by  $n$  individuals. The expected frequency of species with  $n$  individuals is then

$$N_n = S\gamma \frac{x^n}{n} = \alpha \frac{x^n}{n} \quad \dots \dots \dots [14]$$

where  $S$  is the total number of species in the collection and we have put

$$S\gamma = \alpha$$

This expression in equation [14] gives the distribution of species containing different numbers of individuals and it may be written as

$$\alpha x, \frac{\alpha x^2}{2}, \frac{\alpha x^3}{3}, \dots \dots \dots$$

the successive terms being the number of species with 1, 2, 3, ..... individuals per species. This series has an infinite number of terms, is discontinuous and is convergent. The sum of all the terms to infinity gives the total number of species, i.e.,

$$S = -\alpha \log_e(1-x) \quad \dots \dots \dots [15]$$



The corresponding series for the total number of individuals in all species of the same abundance class is

$$\alpha x, \alpha x^2, \alpha x^3, \dots \text{etc.} \dots \dots \dots [16]$$

This is a geometric series with a constant multiple  $x$ . As  $x$  is less than unity the series is convergent, and the sum to infinity ( $N$ ) or the total number of individuals in all the  $S$  species is

$$\begin{aligned} N &= \alpha x + \alpha x^2 + \alpha x^3 + \dots \dots \dots \\ &= x (\alpha + \alpha x + \alpha x^2 + \dots \dots \dots) \end{aligned}$$

The quantity in the bracket is a geometric sequence with  $x$  as quotient whose summation is

$$\left\{ \frac{\alpha}{1-x} \right\} \quad \text{for } |x| < 1$$

$$\therefore N = \frac{\alpha x}{1-x} \dots \dots \dots [17]$$

Equations [15] and [17] lead to

$$\left. \begin{aligned} x &= \frac{N}{N + \alpha} \\ \text{and } S &= \alpha \log_e \left( 1 + \frac{N}{\alpha} \right) \end{aligned} \right\} \dots \dots \dots [18]$$

Now, by knowing the values of the total number of species  $S$  and total number of individuals  $N$  in a particular collection, the values of  $\alpha$  and  $x$  can be

$$577.4 (043)$$

6955

SE

TH-823

calculated using equations [18]. These relationships enable the series to be fitted to any set of observed data. The constant  $\alpha$  has been found to be a measure of the diversity of the population and is low when the number of species is low in relation to the number of individuals, and high when the number of species is high. In other words, when  $\alpha$  is high, abundant species are very much less common than sparse ones, and when  $\alpha$  is low, abundant species are relatively more common.

The magnitude of  $x$  depends only on the size of the sample taken from the parent community. Thus, if we increase the area from which a collection is taken or prolong the duration of operation of a light trap for insects, the only effect will be to change the value of  $x$ , provided the sample still comes from the same parent community.

When Fisher's distribution was fitted to Corbet's uncorrected data of 9031 individuals in 620 species, it gave a value of  $x = 0.997$  and an expected number of species represented by singletons as 135 species, which is considerably more than the observed 118. The calculated and observed values for the difference  $n_1 - n_4$  were 145 and 142, and for  $n_5 - n_{13}$  were 145 and 157. Here  $n_1$  is the number of species represented by singletons

$n_4$  is the number of species represented by 4 individuals each and similarly for  $n_5$  and  $n_{13}$ . Thus the estimated number of species is too high for the rarer, becoming too low for the commoner species [for fuller calculations see Fisher et al., <sup>(7)</sup> 1943, pp. 43].

In a similar way, the extensive data of Williams <sup>(2)</sup> on the capture of about 15,609 Lepidoptera of about 240 species in a light-trap at Rothamsted was also fitted to Fisher's distribution. For both common and rare species the calculated values are very close to the observed values. The calculated number of species with one individual is slightly larger (40.14) than the observed (35). By calculation there should be 116.9 species with 10 or fewer individuals and the observed number was 115. But in the case of Heteroptera with 1414 individuals of 57 species, the fit is not so good as in case of Lepidoptera. The observed number of species represented by only one individual is 18 and the calculated one is less than 12. In general there were rather more of the rarer species than the calculated series indicates and fewer of the commoner species. It is clear from here that the results of Fisher's mathematical theory seem to fit the data of Corbet <sup>(1)</sup> and Williams <sup>(2)</sup> reasonably, but they are not satisfactory

enough to be considered as a conclusive proof of the applicability of the Fisher distribution to observational data in ecosystems.

As opposed to the Fisher distribution which fitted the then existing observational data, Preston<sup>(8)</sup> gave a new suggestion, namely, that the log-normal distribution would be a better fit for the observational data than any other distribution. The fuller details of Preston's approach are discussed in the next Chapter.

CHAPTER - IIIPRESTON'S ANALYSIS

The species abundance data reveal that the collections in all cases contain many rare species and a few abundant ones, although, of course, in terms of numbers of individuals those of the few common species far outnumber those of the many rare species. In any case, in view of the large variation in the numbers of individuals per species, it is convenient to plot these numbers on a logarithmic scale. Preston<sup>(8)</sup> considered the number of species  $n_1, n_2, \dots$ , and in general  $n_r$ , represented respectively by one individual, two individuals, .... and in general  $r$  individuals, i.e., he considered the different species frequencies. He then plotted these  $n_r$  against the corresponding  $r$ , choosing of course, a logarithmic scale for  $r$ . Preston found it convenient to introduce a few other technicalities in his analysis. We review these technicalities now:

Since we are dealing with the relative abundance of species, we frequently come across facts like, say, one species is twice as abundant as another. This led Preston to graduate the abscissa as equal increments in

Table-I: Preston's scheme for grouping species frequencies according to " Octaves "

Column-I Name of the Octave	Column-II Arithmetic grouping of the no. of individuals per species	Column-III Corresponding logarithmic grouping	Column-IV Number of species belonging to that octave (frequency)		
			Half of +	All	+ Half of
A	1 - 2	0 - 1	1	-	2
B	2 - 4	1 - 2	2	3	4
C	4 - 8	2 - 3	4	5, 6, 7	8
D	8 - 16	3 - 4	8	9-15	16
E	16 - 32	4 - 5	16	17-31	32
F	32 - 64	5 - 6	32	33-63	64
G	64 - 128	6 - 7	64	65-127	128
H	128 - 256	7 - 8	128	129-255	256
- - - - - etc - - - - -			- - - - - etc - - -		

the logarithms of the number of individuals representing a species and he used as such increments the 'octave', i.e., the interval in which the representation doubled. In other words the midpoint of each octave is double that of the preceding octave. With this the abscissa became simply a scale of octaves which is equivalent to taking "logarithms to the base 2". The actual details of this grouping are presented in table I.

This table clearly shows the relationship between arithmetical grouping and the corresponding logarithmic grouping. For instance, in the case of octave B, the arithmetic group is 2 to 4 and the corresponding logarithmic group is 1 to 2 [since  $\log_2^2 = 1$  and  $\log_2^4 = 2 \log_2^2 = 2$ ]. The way in which the number of species belonging to each octave is calculated is shown in the same table I, column IV. For instance, if a species is represented by 5, 6 or 7 individuals, it clearly falls in octave C. If a species is represented by 8 individuals, octave C is credited with half a species, and octave D is credited with the other half. In a similar way all other octaves are composed. All species falling in, say, one octave may be thought of as having roughly the same degree of abundance, in comparison with those

falling in any other octave. Note that octaves below A will correspond to fractional numbers and will thus not correspond to any physically observable situations. This is indicated in the plots (see figures 1-6) by drawing the curves in this region by broken lines. Preston<sup>(8)</sup> calls his Y-axis, to the left of which is this broken line, as the "veil line". It should be remembered, however, that doubling the size of the sample will double the number of species belonging to each octave which results in shifting the curve by one octave to the right. In other words, an octave which was hitherto not observable has become observable now. It is thus clear from here that if this process of doubling is continued, all the species that exist in the system will become observable and the corresponding curve will be a true representative of the system. Of course, the log normal nature of this curve does not change as it shifts to the right, the only change being in the constants  $n_0$  and possibly 'a' (introduced below) which will obviously have numerical values characterised by the size of the sample chosen.

In all, Preston<sup>(8)</sup> grouped six different sets of data in this manner and plotted with abscissa as the scale of octaves and the frequency of species (i.e.,



the number of species belonging to a particular octave) as the ordinate. He observed that each curve exhibits a maximum in some octave to the right of the first one and the observed octave frequencies first increase and then decrease. The mode of this set of data is that value which occurs with maximum frequency. Therefore, the octave which corresponds to the maximum in the graph containing maximum number of species is known as the modal octave. For convenience, the plots obtained by Preston are all reproduced at the end of this Chapter in Figures 1 to 6. The figure captions in these figures identify the data plotted.

The main point that emerges from this analysis is that in all cases the data is well fitted by a symmetrical normal curve truncated on the left. The general equation of such a curve is

$$n(R) = n_0 e^{-(aR)^2} \quad . \quad . \quad . \quad . \quad [1]$$

where  $n_0$  is the number of species in the modal octave,  $n(R)$  is the number in an octave distance  $R$  octaves from the modal octave and

$$a^2 = \frac{1}{2\sigma^2} \quad . \quad . \quad . \quad . \quad [2]$$

where  $\sigma$  is the measure of the mean square deviation of the population from its mean value.

We thus have

$$R = \left\{ \log_2 i - \log_2 i^* \right\} \dots \dots \dots [3]$$

where  $i$  are the numbers of individuals, the logarithms to the base 2 of which are plotted along the X-axis and  $i^*$  is the particular value of  $i$  corresponding to the peak of the curve. Preston took for any  $i$  corresponding to any octave the mid-value within that octave.  $i^*$  is thus the mid-value for the model octave. Note that

$$\bar{R}^2 = \left\{ \log_2 i - \log_2 i^* \right\}^2 \dots \dots \dots [4]$$

and is given by

$$\bar{R}^2 = \frac{\int_{-\infty}^{+\infty} n(R) R^2 dR}{\int_{-\infty}^{+\infty} n(R) dR} = \frac{\frac{n_0 \sqrt{\pi}}{2a^3}}{\frac{n_0 \sqrt{\pi}}{a}} = \frac{1}{2a^2} \dots [5]$$

$$\equiv \sigma^2.$$

Note also that the total number of species  $N$  for the system is given by

$$\int_{-\infty}^{+\infty} n(R) dR = \frac{n_0 \sqrt{\pi}}{a} = N \dots \dots \dots [6]$$

Table-II: Calculated values of  $a$ ,  $n_0$ ,  $N$  and the observed value of the total number of species for cases analysed by Preston.

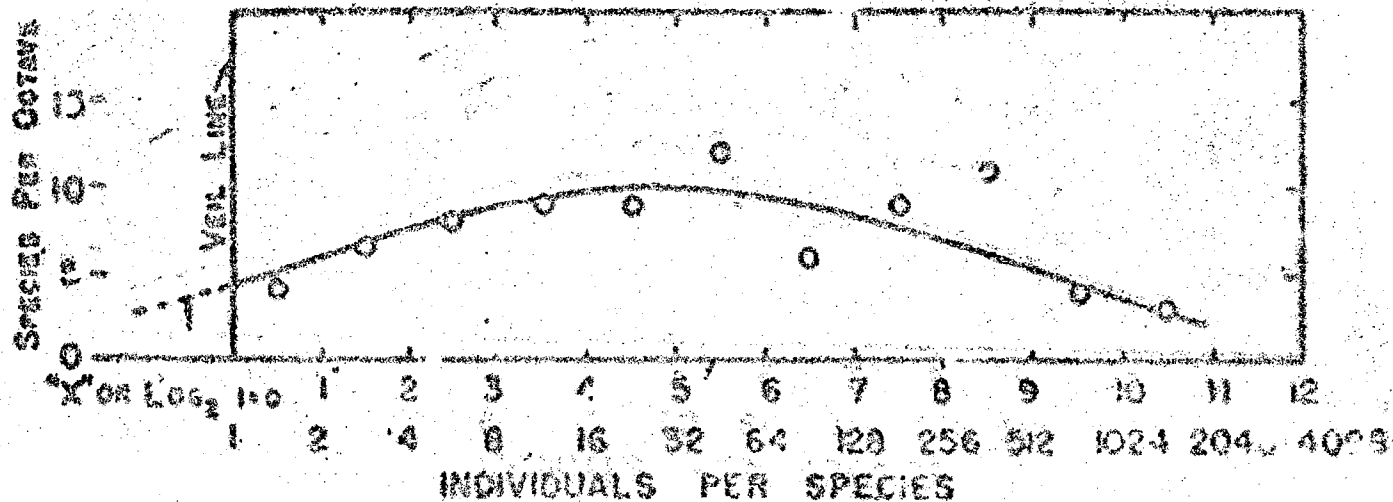
Case	$a$	$n_0$	$N$	Observed total number of species in the particular sample analysed.
1	0.194	10	91	80
2	0.207	48	410	349
3	0.205	42	363	226
4	0.227	35	273	240
5	0.152	33	384	277
6	0.160	30	332	291

This  $N$  should be the total number theoretically available for observation. In practice, the presence of the veil line will imply that the actually observed number is smaller. Now since the quantities  $a$  and  $n_0$  can be calculated from the curves fitted to the data, the theoretical value of the total number of species for the system can be calculated from equation [6]. The difference between this and the observed value will give an estimate of how much of the system was missed out in any particular sample. In table II, we present the calculated values of  $a$ ,  $n_0$ ,  $N$  as well as the observed values (in the given samples) for  $N$ , for all the six cases analysed by Preston. Note that the value of  $a$  in all these cases is close to about 0.20.

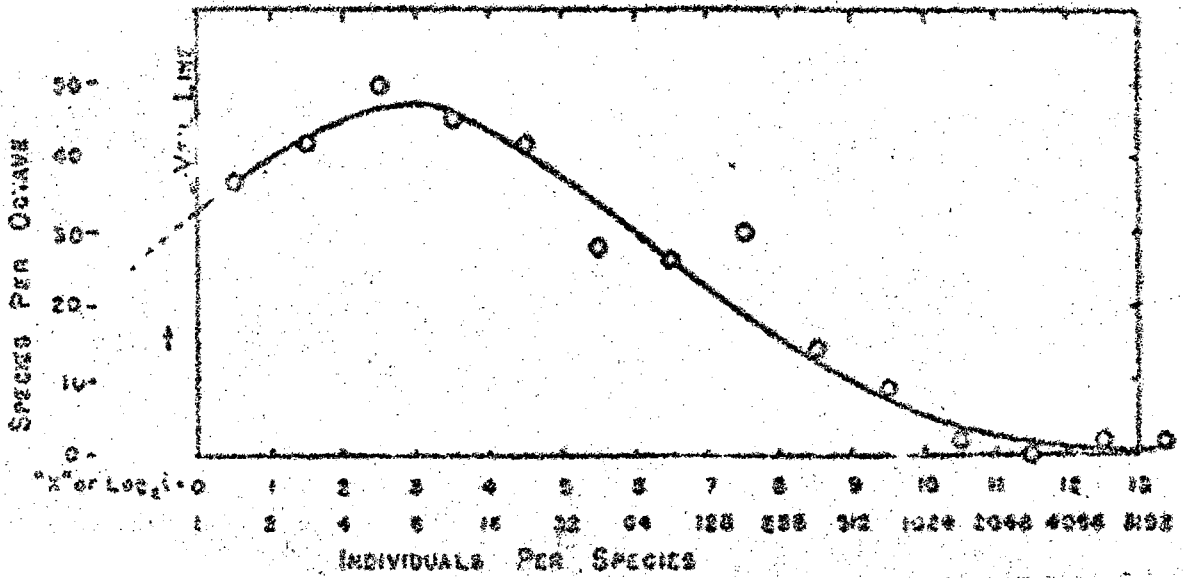
We conclude this Chapter by noting that the distribution given by equation [1] can be conveniently written in the form

$$n(\log_2 i) = \frac{N}{\sigma \sqrt{2\pi}} e^{-\left\{ \frac{\log_2 i - \log_2 i^*}{2\sigma^2} \right\}^2} \dots [7]$$

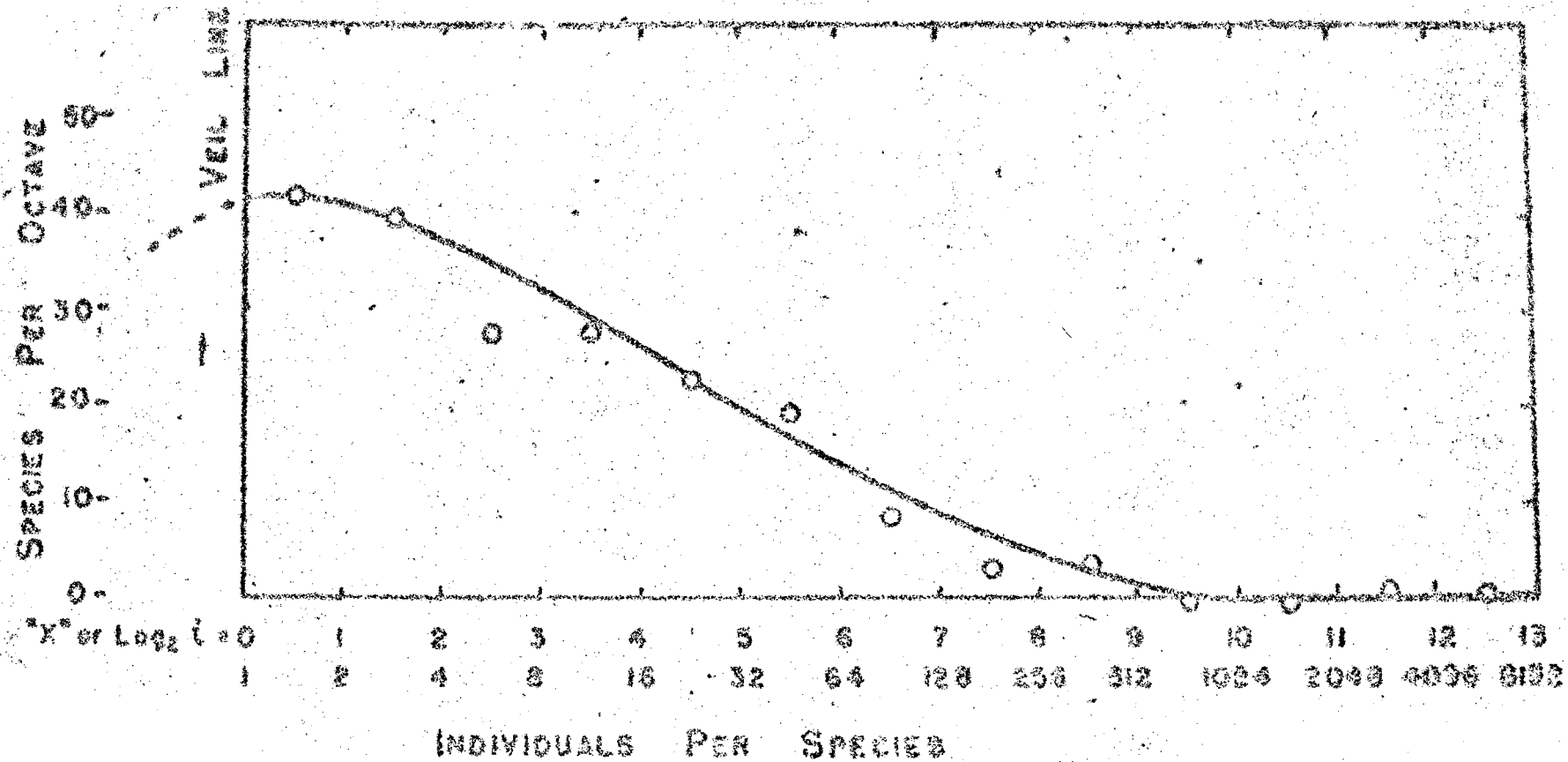
which, aside from the trivial difference of the logarithms being taken to the base '2' instead of the base 'e' is identical to the standard lognormal distribution. We shall return to this expression in the last Chapter of this thesis.



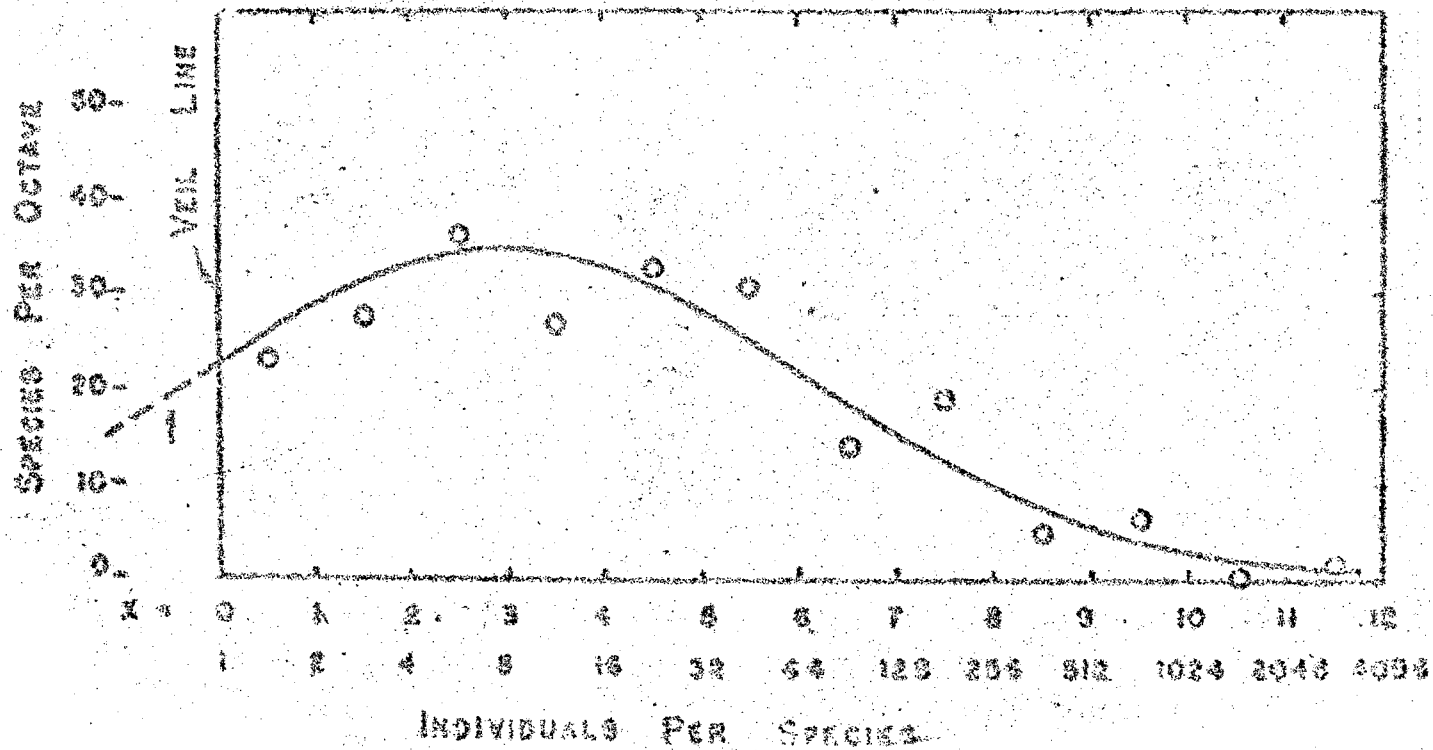
**Figure 1:** Species abundances in a collection of birds of Quaker Run Valley, Western New York State. Data from Saunders (3) (1936).



**Figure 2:** Species abundance in a collection of moths caught in a light trap at Orono, Maine. Data from Dirks<sup>(5)</sup> (1937).

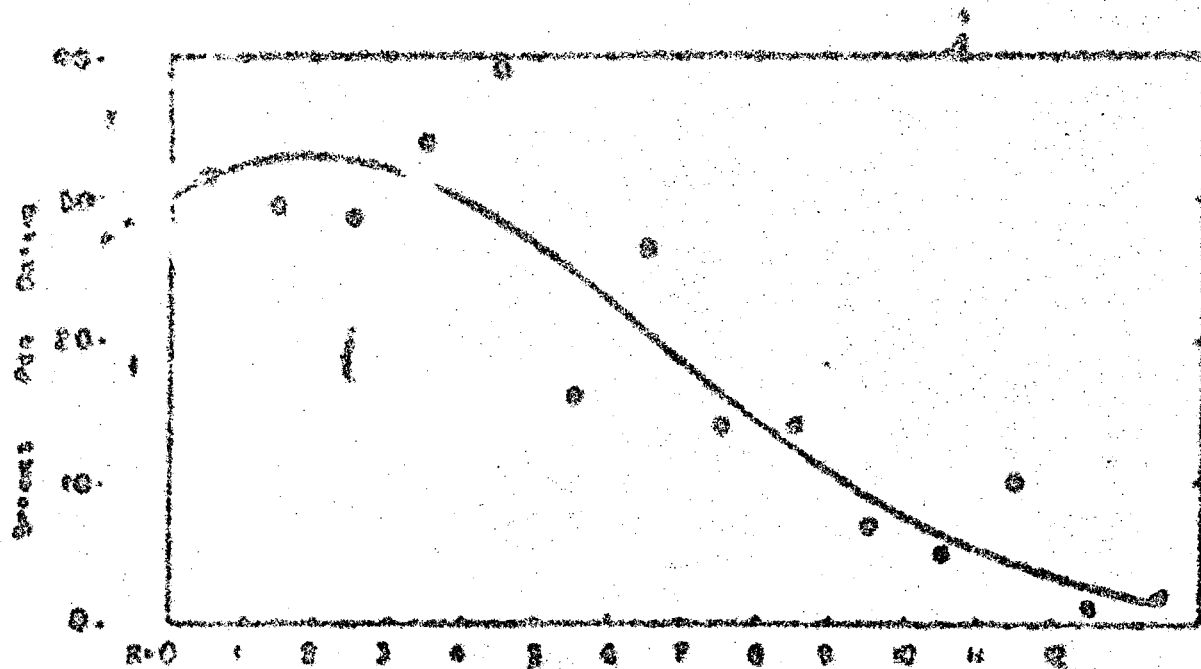


**Figure 3:** Species abundances in a collection of female moths in a light trap at Orono, Maine.  
Data from Dirks<sup>(5)</sup> (2937).

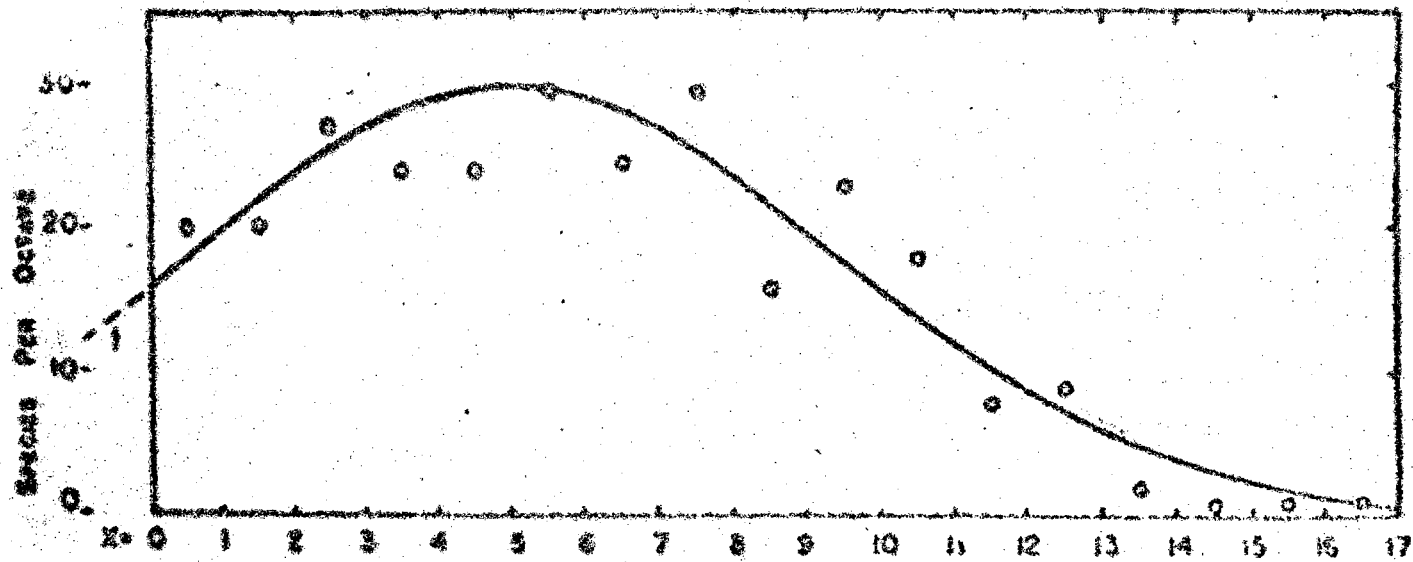


**Figure 4:** Species abundances in a collection of moths in a light trap at Rothamsted, England.  
Data from Williams<sup>(2)</sup> (1943).





**Figure 5:** Species abundances in a collection of moths caught in a light trap at Saskatoon, Saskatchewan.  
Data from King.



**Figure 6:** Species abundances in a collection of moths caught in a light trap at Lethbridge, Alberta. Data from Seamans (6).

CHAPTER - IVSTATISTICAL MECHANICAL APPROACH TO THE  
LOTKA-VOLTERRA MODEL

In this Chapter, we shall develop the statistical mechanics of the multispecies ecosystem, the dynamics of which is given by the coupled Lotka-Volterra equations. We shall see that, under a constraint, this set of equations possesses a constant of motion. We then construct a phase space, such that our system belongs to an ensemble in this space on a surface characterised by this constant of motion. The requisite Liouville theorem, the condition for the ensemble to be in statistical equilibrium, as well as the ergodic hypothesis are then discussed. Following this, microcanonical ensemble is constructed. The possibility of there being a "temperature" for the system is then noted. This enables us to construct the Gibbs canonical ensemble for any subsystem of the system. As an application, the probability distribution for the number of individuals in any species is then worked out. This distribution is seen to be the  $\chi^2$  distribution used by Fisher and discussed in the Introduction and the second Chapter of this thesis.

1. THE LOTKA-VOLTERRA MODEL

The basic equations of the Lotka-Volterra model for a system of  $n$  interacting species with populations  $N_1, N_2, \dots, N_n$  are:

$$\frac{dN_r}{dt} = \epsilon_r N_r + \frac{1}{\beta_r} \sum_{S=1}^n \alpha_{Sr} N_S N_r \dots [1]$$

where  $S, r = 1, 2, 3, \dots, n$ .

The first term involving the constants  $\epsilon_r$  on the right hand side here expresses how the species  $N_r$  propagates if left to itself in a given environment. The remaining terms express the interaction of species  $r$  with all other species. The constants  $\alpha_{Sr}$  are assumed to be antisymmetric with

$$\alpha_{Sr} = -\alpha_{rS} \quad \dots \quad [2]$$

This of course implies that the self-interaction terms are absent, i.e.,

$$\alpha_{rr} = 0, \quad \text{for all } r \quad \dots \quad [3]$$

The quantities  $\beta_r^{-1}$  are such that during binary collisions of species  $r$  and  $S$ , the ratio of  $S$ 's lost (or gained) per second to  $r$ 's gained (or lost) is  $\beta_S^{-1}/\beta_r^{-1}$ .

We define the steady state of our assembly to be characterized by that set of populations  $N_r \equiv q_r$

for which  $\frac{dN_r}{dt} = 0$  for all  $r$ , so that the defining equations for  $q_r$  are

$$\epsilon_r \beta_r + \sum_S \alpha_{Sr} q_S = 0 \quad . . . . [4]$$

It is assumed that at steady state conditions, none of the  $q$  values will vanish. Now we define

$$v_r \equiv \log \frac{N_r}{q_r} \quad \text{or} \quad N_r = q_r e^{v_r} \quad . . . [5]$$

Clearly, as  $N_r \rightarrow q_r$ ,  $v_r \rightarrow 0$  so that  $v_r$  is a measure of the deviation from equilibrium. The rate equation [1] may be expressed in terms of the  $v_r$  as :

$$\beta_r \frac{dv_r}{dt} = \epsilon_r \beta_r + \sum_S \alpha_{Sr} q_S e^{v_S} \quad \dots [6]$$

and by expressing  $\epsilon_r \beta_r$  through equation [4] we have

$$\beta_r \frac{dv_r}{dt} = \sum_S \alpha_{Sr} q_S \{e^{v_S} - 1\} \quad . . . . [7]$$

If we multiply both sides of this equation by  $q_r \{e^{v_r} - 1\}$  and sum over all  $r$ , we find

$$\begin{aligned} & \frac{d}{dt} \left\{ \sum_r \beta_r q_r \left\{ e^{v_r} - v_r \right\} \right\} \\ & = \sum_{rs} \alpha_{sr} q_s q_r \left\{ e^{v_s - 1} \right\} \left\{ e^{v_r - 1} \right\} = 0 \end{aligned}$$

the double sum vanishes because  $\alpha_{sr}$  are antisymmetrical in  $s$  and  $r$ . Then, we find

$$G = \sum_{\mathbf{r}} \beta_{\mathbf{r}} q_{\mathbf{r}} \left\{ e^{v_{\mathbf{r}}} - v_{\mathbf{r}} \right\} = \text{Cont} = \sum G_{\mathbf{r}} \dots [8]$$

so that  $G$  is our desired constant of motion.

Now the equations of motion [7] may be written canonically in terms of  $G$  as follows :

$$\dot{v}_{\mathbf{r}} = \sum_{\mathbf{s}} \frac{\alpha_{\mathbf{s}\mathbf{r}} q_{\mathbf{s}}}{\beta_{\mathbf{r}}} \left\{ e^{v_{\mathbf{s}}} - 1 \right\}$$

$$\dot{v}_{\mathbf{r}} = \left\{ \sum_{\mathbf{s}} \gamma_{\mathbf{s}\mathbf{r}} \frac{\partial}{\partial v_{\mathbf{s}}} \right\} G \dots \dots \dots [9]$$

where  $\left\{ \gamma_{\mathbf{s}\mathbf{r}} = \frac{\alpha_{\mathbf{s}\mathbf{r}}}{\beta_{\mathbf{s}} \beta_{\mathbf{r}}} = - \gamma_{\mathbf{r}\mathbf{s}} \right\}$

Equations [9] are the canonical forms of [7] in terms of  $G$ .

The fact that  $G$  as given by [8] is a sum of individual terms, each relating to a separate species is of considerable importance. Each  $G_{\mathbf{r}}$  in [8] has a minimum  $\beta_{\mathbf{r}} q_{\mathbf{r}}$  ( $\equiv \gamma_{\mathbf{r}}$ ) when  $v_{\mathbf{r}} = 0$  or when  $N_{\mathbf{r}} = q_{\mathbf{r}}$ , the equilibrium value of  $N_{\mathbf{r}}$ . To say that  $G$  is a constant of the motion implies that if the system is started off with  $G = G_0$ , then the value  $G_0$  will be maintained indefinitely. This result  $G = G_0$  may also be used, as will be done later, to define a constant surface in

the phase space for the system described by the set of equations [1].

Kerner<sup>(10)</sup> built the statistical mechanics of the system [1] by imagining a Gibbs ensemble for it in the following way. He considered all possible copies of the system controlled by the dynamical equation [1], where each such copy corresponded to one of the possible sets of initial values of  $(v_r)$  consistent with the constant  $G$ . The state of each such copy was represented by a point in the cartesian space of the variables  $(v_1, v_2, \dots, v_n)$ , called the phase space. The collection of all such points then defined the ensemble.

## 2. THE LIOUVILLE THEOREM AND THE CONDITION FOR STATISTICAL EQUILIBRIUM

Now, let us denote the density of the points in the ensemble by  $\rho(v_1, \dots, v_n)$  at the point  $(v_1, \dots, v_n)$  in the phase space. Clearly, the total number of copies of the system, i.e., the total number of points in the ensemble, has to be a constant. This conservation property immediately leads to the equation of continuity for  $\rho(v_1, \dots, v_n)$  which is:

$$\frac{\partial P(v_1, \dots, v_n)}{\partial t} + \sum_{\mathbf{r}} \frac{d \dot{v}_{\mathbf{r}}}{d v_{\mathbf{r}}} P(v_1, \dots, v_n) = 0 \dots \dots [10]$$

Expanding the derivative under the summation sign, we get

$$\frac{\partial P}{\partial t} + \sum_{\mathbf{r}} \dot{v}_{\mathbf{r}} \frac{\partial P}{\partial v_{\mathbf{r}}} + \sum_{\mathbf{r}} P \frac{\partial \dot{v}_{\mathbf{r}}}{\partial v_{\mathbf{r}}} = 0 \dots \dots [11]$$

In view of equation [7], which implies that  $\dot{v}_{\mathbf{r}}$  is independent of  $v_{\mathbf{r}}$ , we have

$$\frac{\partial P}{\partial t} + \sum_{\mathbf{r}} \frac{\partial P}{\partial v_{\mathbf{r}}} \frac{\partial v_{\mathbf{r}}}{\partial t} = 0 \dots \dots [12]$$

which is simply the relation

$$\frac{DP}{Dt} = 0 \dots \dots [13]$$

We have thus proved the Liouville theorem.

This implies that a given element of volume of phase space, though changing it's shape, maintains a uniform size, as the motion of its points unfolds, thus giving the desirable uniformity property to the phase space.

Let us now consider our actual system which is characterized by the constant  $G = G_0$ . The ensemble for the system will thus be an ensemble over a surface of constant  $G$  in the phase space. We then presume the



density function  $P$  to be a function of  $G$  alone. The dependence of  $P$  on any other variable is thus only via  $G$ . Then we have

$$\frac{\partial P}{\partial v_r} = \frac{\partial P}{\partial G} \frac{\partial G}{\partial v_r} \quad . \quad . \quad . \quad . \quad . \quad . \quad [14]$$

and substituting this in the second term in equation [12], we get

$$\begin{aligned} \sum_r \frac{\partial P}{\partial v_r} \cdot \frac{\partial v_r}{\partial t} &= \frac{\partial P}{\partial G} \sum_r \frac{\partial G}{\partial v_r} \frac{\partial v_r}{\partial t} \\ &= \frac{\partial P}{\partial G} \left\{ \frac{dG}{dt} \right\} \\ &= 0 \quad . \quad . \quad . \quad . \quad [15] \end{aligned}$$

in view of the constancy of  $G$ .

This immediately gives us the result

$$\frac{\partial P}{\partial t} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad [16]$$

which we obtain by substituting equation [15] in equation [12]. This is the condition for our ensemble to be in statistical equilibrium.

In order to proceed further with our system in a situation of statistical equilibrium, we parametrise our phase density as

$$P = P_0 \delta (G - G_0) \quad . \quad . \quad . \quad . \quad . \quad [17]$$

where  $\delta$  stands for the standard Dirac delta function and  $P_0$  is a numerical constant.  $G_0$  is the initial value of  $G$ , defining a constant surface in the phase space. The ensembles characterized with density functions of this kind are the well known microcanonical ensembles.

The ensemble average of any function  $f(v_1, v_2, \dots, v_n)$  of phase coordinates is now defined to be

$$f = \frac{\int_{-\infty}^{+\infty} P f d\Gamma}{\int_{-\infty}^{+\infty} P d\Gamma} \quad \dots \quad [18]$$

the integrals being over all of phase space.

The element of volume  $d\Gamma$  can be represented as

$$d\Gamma = ds dn = ds \frac{dG}{|\nabla G|}$$

where  $ds$  is an element of area on a constant  $G$  surface and  $dn$  is an increment of length normal to the surface, which can be written as  $\frac{dG}{|\nabla G|}$  where  $dG$  is the difference in  $G$  values two neighbouring constant  $G$  surfaces. Therefore, with this result and the equation [17], the expression [18] goes over to

$$\begin{aligned} \bar{f} &= \frac{\int P_0 f \delta (G - G_0) ds \frac{dG}{|\nabla G|}}{\int P_0 \delta (G - G_0) ds \frac{dG}{|\nabla G|}} \\ &= \frac{\int_{G_0} f \frac{ds}{|\nabla G|}}{\int_{G_0} \frac{ds}{|\nabla G|}} \dots \dots \dots [19] \end{aligned}$$

Note that these integrals are surface integrals over  $G = G_0$ .

Let us now choose for  $f$ , the function  $T_r$  given by

$$T_r = v_r \frac{\partial G}{\partial v_r} = \gamma_r \left\{ \frac{N_r}{q_r} - 1 \right\} \log \frac{N_r}{q_r} ,$$

where,  $\gamma_r = \beta_r q_r$ .

We note first that since

$$\nabla G = \sum \frac{\partial G}{\partial v_r} \hat{v}_r = |\nabla G| \hat{n}$$

( $\hat{v}_r$  denoting a unit vector in the  $v_r$  direction and  $\hat{n}$  a unit normal vector to the surface  $G = \text{constant}$ ), the direction cosines of  $\hat{n}$  are

$$\hat{n} \cdot \hat{v}_r = \frac{\partial G}{\partial v_r} / |\nabla G|$$

so that

$$\begin{aligned} v_r \frac{\partial G}{\partial v_r} \frac{ds}{|\nabla G|} &= v_r (\hat{n} \cdot \hat{v}_r) ds \\ &= \hat{n} \cdot \vec{v}_r ds. \end{aligned}$$

Here,  $\vec{v}_r$  denotes the vector to the point  $(0, \dots, 0, v_r, 0, \dots, 0)$ . Then calling the denominator in equation [19] as  $\Delta_0$ , the ensemble average of  $T_r$  is

$$\begin{aligned} \bar{T} &= \frac{1}{\Delta_0} \int \hat{n} \cdot v_r ds = \frac{1}{\Delta_0} \int \text{div } v_r d\tau \\ &= \frac{\gamma_0}{v_0}, \text{ for all } \gamma \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad [20] \end{aligned}$$

where  $\gamma_0$  denotes the total volume enclosed by  $G_0$ . Note here that we have used Gauss' divergence theorem. This expression  $\frac{\gamma_0}{\Delta_0}$  in equation [20] is a constant for all values of  $\gamma$  (i.e., for all species). Thus the mean  $T_r$  for any species is the same as for any other.

This result is analogous to the equipartition theorem which states that the energy of a system of particles is equally distributed among all the degrees of freedom of the particles in the system. It is  $\frac{1}{2}$  KT per degree of freedom, where K is the Boltzmann constant and T is the absolute temperature of the system. Since the idea of equipartition prevails in our biological system as well, as evidenced by the properties of  $\bar{T}_r$ , the possibility of defining a 'temperature' for the system opens up immediately. Once a temperature is defined, a cononical ensemble can also be

defined. We shall do precisely this in the following, but first we shall show by an example that our system admits of the ergodic hypothesis as well.

### 3. THE ERGODIC HYPOTHESIS

We define a function  $D_r$  as

$$D_r = \frac{\partial G}{\partial v_r} = \gamma_r \left\{ \frac{N_r}{q_r} - 1 \right\} \dots \dots \dots [21]$$

Let us calculate the ensemble average of  $D_r$ , i.e.,

$$\begin{aligned} \bar{D}_r &= \frac{\int_{G_0} \frac{\partial G}{\partial v_r} \frac{ds}{|\nabla G|}}{\Delta_0} \\ &= \frac{\int_{s_0} (\hat{n} \cdot \hat{v}_r) ds}{\Delta_0} \\ &= \frac{1}{\Delta_0} \int \text{div } \hat{v}_r d\mathcal{L} \\ &= 0. \end{aligned}$$

Since the divergence of a unit vector is zero.

$$\therefore \bar{D}_r = \gamma_r \left\{ \frac{\bar{N}_r}{q_r} - 1 \right\} = 0 \dots \dots \dots [22]$$

$$\text{or, } \bar{N}_r = q_r \dots \dots \dots [23]$$

In other words the ensemble average of  $N_r$  is equal to the steady state value of  $N_r$ , namely  $q_r$ . We can now show that the time average of  $N_r$  is also  $q_r$ . We have from equation [1],

$$\beta_r \frac{d}{dt} \left\{ \log N_r \right\} = \epsilon_r \beta_r + \sum_s \alpha_{sr} N_s,$$

Integrating with respect to  $t$  on both sides, between the limits 0 to  $T$ , we get,

$$\frac{\beta_r}{T} \log \left\{ \frac{N_r(T)}{N_r(0)} \right\} = \epsilon_r \beta_r + \sum_s \alpha_{sr} \left\{ \frac{1}{T} \int_0^T N_s(t) dt \right\}$$

Now, as  $T \rightarrow \infty$ , the left hand side vanishes and remembering that the time average of  $N_s$  is nothing but,

$$\bar{N}_s = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T N_s(t) dt, \text{ we simply get}$$

$$\sum_s \alpha_{sr} \bar{N}_s = - \epsilon_r \beta_r$$

By comparing this with equation [4], we arrive at the result

$$\bar{N}_s = q_s \quad . \quad . \quad . \quad . \quad . \quad . \quad [24]$$

Thus, by combining equations [23] and [24], we see that the time average of population  $N_r$  is precisely equal to the ensemble average of the same. The ergodic hypothesis, first advanced by Boltzmann, states that a

time average of a macroscopic quantity is, under equilibrium conditions, the same as an ensemble average. It is clear from our result that our biological system is ergodic.

#### 4. THE CANONICAL ENSEMBLE

Our biological system consists of  $n$  species in general. If we were to consider subsystems of this general system, which would consist of say,  $\nu$  number of species, where  $1 \leq \nu \leq n$ , it would be convenient to work with the Gibbs canonical ensemble instead of the microcanonical ensemble that was discussed in the preceding paragraphs. The canonical ensemble can of course be defined for any subsystem which is in equilibrium with the larger system of which it is a part. This equilibrium is established once we identify the temperature for the system and ensure that the temperature of the subsystem is the same as that of the full system. The existence of the quantity  $\bar{T}_R$ , which is the same for any  $\nu$ , makes it clear that we can indeed define a 'temperature' in our case as well. If we now denote this temperature by the symbol  $\Theta$ , and consider the subsystem consisting of  $\nu$  number of species, we can construct our canonical ensemble which is characterized by the density function

$$P_{\nu} = \frac{e^{(-G_{\nu}/\theta)}}{\int e^{(-G_{\nu}/\theta)} d\mathcal{L}_{\nu}}$$

where,  $P_{\nu}$  is normalised:

$$\int P_{\nu} d\mathcal{L}_{\nu} = 1.$$

$P_{\nu} (v_1, v_2, \dots, v_{\nu}) d\mathcal{L}_{\nu}$  here represents the probability that a member of the ensemble (which is in statistical equilibrium) chosen at random will be found in the volume element  $d\mathcal{L}_{\nu}$  around the point  $(v_1, v_2, \dots, v_{\nu})$  in the  $\nu$ -dimensional subspace of our subsystem. The canonical average of any function  $f$  is now given by

$$\langle f \rangle \equiv \bar{f} \equiv \int f P_{\nu} d\mathcal{L}_{\nu}.$$

Note that the factor

$$Z \equiv \int e^{-G_{\nu}/\theta} d\mathcal{L}_{\nu} \quad . \quad . \quad . \quad . \quad [25]$$

is the well known Gibbs phase integral.

We now calculate the canonical average of the function  $D_{\mathbf{r}}$ , defined in equation [21]. We have:

$$\bar{D}_{\mathbf{r}} = \frac{\int \frac{\partial G}{\partial v_{\mathbf{r}}} e^{-G/\theta} d\mathcal{L}}{\int e^{-G/\theta} d\mathcal{L}} \quad . \quad . \quad . \quad . \quad [26]$$

where for convenience, we have dropped the subscript  $\nu$ .

By making the substitution,



$$\frac{\partial}{\partial v_r} \left\{ e^{-G_r/\theta} \right\} = -\frac{1}{\theta} e^{-G_r/\theta} \left\{ \frac{dG}{dv_r} \right\}$$

into equation [26], we have

$$\bar{D}_r = \frac{\int \frac{\partial G}{\partial v_r} \left\{ e^{-G_r/\theta} \right\} e^{-\frac{G_{rest}}{\theta}} d\Gamma}{\int e^{-G_r/\theta} e^{-G_{rest}/\theta} d\Gamma}$$

where,  $G = G_r + G_{rest}$

$$\text{or, } \bar{D}_r = \frac{-\theta \int_{-\infty}^{+\infty} \frac{\partial}{\partial v_r} \left\{ e^{-G_r/\theta} \right\} e^{-G_{rest}/\theta} dv_r dv_{rest}}{\int_{-\infty}^{+\infty} e^{-G_r/\theta} e^{-G_{rest}/\theta} dv_r dv_{rest}}$$

$$= -\theta \left[ e^{-G_r/\theta} \right]_{-\infty}^{+\infty} / \int_{-\infty}^{+\infty} e^{-G_r/\theta} dv_r$$

$$= 0 \quad . \quad . \quad . \quad . \quad . \quad [27]$$

Therefore, as for microcanonical ensemble, we get for canonical ensemble :

$$\gamma_r \left\{ \frac{\bar{N}_r}{q_r} - 1 \right\} = 0$$

$$\text{or } \bar{N}_r = q_r.$$



Here we make the substitution,

$$\frac{\partial}{\partial v_r} \left\{ \frac{\partial G_r}{\partial v_r} e^{-G_r/\theta} \right\} = -\frac{1}{\theta} \left\{ \frac{\partial G_r}{\partial v_r} \right\}^2 e^{-G_r/\theta} + \frac{\partial^2 G_r}{\partial v_r^2} e^{-G_r/\theta},$$

$$\text{where, } G_r = \gamma_r ( e^{v_r} - v_r )$$

$$\frac{\partial G_r}{\partial v_r} = \gamma_r ( e^{v_r} - 1 ),$$

$$\text{and } \frac{\partial^2 G_r}{\partial v_r^2} = \frac{\partial G_r}{\partial v_r} + \gamma_r .$$

With these substitutions, we have

$$\begin{aligned} \int \left\{ \frac{\partial G_r}{\partial v_r} \right\}^2 e^{-G_r/\theta} dv_r &= -\theta \left\{ \frac{\partial G_r}{\partial v_r} e^{-G_r/\theta} \right\}_{-\infty}^{+\infty} + \theta \int \left\{ \frac{\partial G_r}{\partial v_r} \right\} e^{-G_r/\theta} dv_r \\ &\quad + \theta \gamma_r \int e^{-G_r/\theta} dv_r . \end{aligned}$$

The first and second integrals in the right hand side of the above equation are zero. Therefore, we have,

$$\int \left\{ \frac{\partial G_r}{\partial v_r} \right\}^2 e^{-G_r/\theta} dv_r = \theta \gamma_r \int e^{-G_r/\theta} dv_r \dots [30]$$

With this expression [30], equation [29] becomes

$$\begin{aligned} \bar{D}_r^2 &= \frac{\Theta \gamma_r \int e^{-G_r/\Theta} dv_r}{\int e^{-G_r/\Theta} dv_r} \\ &= \Theta \gamma_r = \gamma_r^2 \left\{ \frac{N_r}{q_r} \right\}^2. \end{aligned}$$

Therefore,

$$\begin{aligned} \Theta &= \gamma_r \left\{ \frac{N_r}{q_r} - 1 \right\}^2 \\ &= \frac{\beta_r}{q_r} \left\{ N_r - q_r \right\}^2 \dots \dots \dots [31] \end{aligned}$$

Thus,  $\Theta$  is a measure of the mean square deviation in the numbers for each species from its equilibrium value. The situation of zero temperature would thus correspond to the completely quiet state of the biological system. Clearly,  $\Theta$  is a measure of the level of excitation of the system from its stationary state.

##### 5. KERNER DISTRIBUTION

We now consider the Gibbs phase integral set out in equation [25] to obtain an expression for the probability that the species  $\gamma$  will have its  $v_r$  in the range  $v_r$  and  $v_r + dv_r$ . We have

$$Z = \int e^{(-\alpha \sum_r G_r)} d\tau$$

$$= \prod \int_{-\infty}^{+\infty} e^{-\alpha G_r} dv_r$$

where  $\alpha \equiv \frac{1}{\theta}$  and  $\prod$  denotes the product symbol.

$$\text{Since } e^{-\alpha G_r} = e^{-\alpha \gamma_r (e^{v_r} - v_r)},$$

$$= e^{-\alpha \gamma_r \left\{ \frac{N_r}{q_r} \right\}} \left\{ \frac{N_r}{q_r} \right\}^{\alpha \gamma_r},$$

the expression for Z becomes

$$Z = \prod \int_{-\infty}^{+\infty} e^{-\alpha \gamma_r \left\{ \frac{N_r}{q_r} \right\}} \left\{ \frac{N_r}{q_r} \right\}^{\alpha \gamma_r} \frac{dN_r}{N_r},$$

$$= \prod (\alpha \gamma_r)^{-\alpha \gamma_r} \Gamma(\alpha \gamma_r)$$

$$= \prod Z_r$$

where,  $Z_r = (\alpha \gamma_r)^{-\alpha \gamma_r} \Gamma(\alpha \gamma_r) \dots \dots \dots [32]$

with this expression, the probability that one species will have its  $v_r$  in the range  $v_r$  and  $v_r + dv_r$  is given by

$$P_r dv_r = \frac{e^{-G_r/\theta} dv_r}{Z_r} \dots \dots \dots [33]$$

Transforming the variables  $v_r$  into a new variable  $n_r$  such that  $n_r = \frac{N_r}{q_r}$ , we have,

$$\begin{aligned} e^{-G_r/\theta} &= e^{-\alpha\gamma_r n_r} e^{\alpha\gamma_r \log n_r} \\ &= e^{-\alpha\gamma_r n_r} (n_r)^{\alpha\gamma_r} \dots [34] \end{aligned}$$

With expressions [32] and [34], the probability that the species will have its  $n_r$  in the range  $n_r$  and  $n_r + dn_r$  is given by

$$Pn_r dn_r = \frac{(n_r)^{\alpha\gamma_r - 1} e^{-\alpha\gamma_r n_r} dn_r}{(\alpha\gamma_r)^{-\alpha\gamma_r} \Gamma(\alpha\gamma_r)} \dots [35]$$

This expression [35] is the  $\chi^2$  distribution chosen by Fisher<sup>(7)</sup> for the intrinsic abundance of species - the expectation value of the poisson distribution, as discussed in the second Chapter. Fisher chose this  $\chi^2$  distribution only at a phenomenological level and the resulting compound poisson distribution which he got from this led to a negative binomial distribution used by him to fit the observational data. The result of Kerner thus gives a theoretical reasoning as to why the  $\chi^2$  distribution should be used.

However, as has been noted in the introduction, the distribution which is actually the more extensively valid one, is not the Fisher distribution, but rather the standard lognormal distribution suggested by Preston<sup>(8)</sup> (discussed in Chapter III). Interestingly enough, we find that if Kerner's approach is followed within a model where interactions between species are of the Gompertz form, a distribution for  $N_i$  follows which is precisely lognormal. We show this in the next Chapter.

CHAPTER - VSTATISTICAL MECHANICAL APPROACH TO THE GOMPERTZ MODEL

In the previous Chapter, we have discussed the application of statistical mechanical techniques to the Lotka-volterra model. These techniques can of course be tried on other models of multispecies ecosystems also. We shall do so in the present Chapter, our model will differ from the Lotka-volterra model insofar as the interaction terms between different species are concerned. Consider for instance the interaction term appearing in the equation for the time derivative of population  $N_i$  of the  $i^{\text{th}}$  species, and representing the interaction between the species  $i$  and, say,  $j$ . In the Lotka-volterra model it is simply of the form  $N_i N_j$ . In the present model, it will be taken as  $N_i \log N_j$ . Models with such interaction forms, commonly called the Gompertz forms have been discussed by several authors. See for instance Bhat and Pande (1980)<sup>11</sup>, we give a detailed discussion on 2 and 3 species ecosystems within such a model. This model reproduces all the good features of the conventional Lotka-volterra model. In addition, because of its solvability property, it is amenable to much more, detailed as well as fruitful investigations than any other



model, atleast for cases where the number of interacting species is low enough. We have thought it quite useful to try this model for cases where the number of interacting species is large. The statistical mechanical techniques for such cases are of course the appropriate ones.

We shall see in the following that our model leads to the log normal form for the probability distribution for the number of individuals in any species, under an approximation, which is quite an acceptable one so long as we confine ourselves to species within one genera, we shall show that, this leads directly to an explanation of Preston result discussed in Chapter III.

#### 1. STATISTICAL MECHANICS OF THE GOMPertz MODEL

We now consider the model of  $n$  interacting species where the interaction term appearing in the equation for the time derivative of  $N_r$ , and representing the interaction between species  $r$  and  $s$ , is of the form  $N_r \log N_s$ . The basic equations of the model are thus

$$\frac{dN_r}{dt} = \epsilon_r N_r + \sum_s \alpha_{sr} N_r \log N_s \dots \dots [1]$$

$$r, s = 1, 2, \dots n.$$

Once again,  $\alpha_{sr}$  are assumed to be antisymmetric. For stationary states of the system,  $\frac{dN_r}{dt}$  vanishes and the population numbers  $N_r$  have the steady values  $q_r$  given by

$$\epsilon_r q_r + \sum_s \alpha_{sr} q_r \log q_s = 0 \dots \dots \dots [2]$$

Now in terms of the variable  $x_r$ , given by

$$x_r = \log N_r \dots \dots \dots [3]$$

equations [1] can be written as

$$\frac{dx_r}{dt} = \epsilon_r + \sum_s \alpha_{sr} x_s \dots \dots \dots [4]$$

and by expressing  $\epsilon_r$  through equation [2], we have

$$\dot{x}_r = \sum_s \alpha_{sr} (x_s - x_s^*) \dots \dots \dots [5]$$

where  $x_s^* = \log q_s$ .

Multiplying both sides of equation [5] by  $(x_r - x_r^*)$  and summing over  $r$ , we get

$$\sum_r \dot{x}_r (x_r - x_r^*) = \sum_{sr} \alpha_{sr} (x_r - x_r^*) (x_s - x_s^*)$$

Due to the antisymmetry of  $\alpha_{sr}$ , the right hand side of the above equation vanishes and the equation yields a constant of motion or a conserved quantity, given by

$$G = \sum_r \left\{ \frac{1}{2} x_r^2 - x_r x_r^* \right\} \equiv \sum_r G_r \dots \dots [6]$$

We now imagine a Gibbs ensemble for the system [4], in a manner analogous to that of Kerner, as the collection of all possible copies of the system, each copy corresponding to one possible set of initial values for  $x_r$ . An appropriate phase space - a cartesian coordinate system with  $x_r$  as the axes, can now be constructed. Each copy of the system is then represented in this space by a point, and the ensemble by an ensemble of points. The constancy of the total number of points in the ensemble, along with the dynamics of our model as given by equations [1] and [4], then leads to the requisite Liouville theorem. The ensemble for our system is of course defined over a surface of constant  $G$  in the phase space. The condition for this ensemble to be in a state of statistical equilibrium can also be established in exactly the same way as in the previous Chapter. All this enables us to define a microcanonical ensemble for our system, the density function for which is characterised by :

$$P = P_0 \delta (G - G_0) \cdot \cdot \cdot \cdot [7]$$

where, as before,  $\delta$  stands for the standard Dirac delta function and  $P_0$  is a numerical constant.

The ensemble average of any function  $f(x_1, x_2, \dots, x_n)$  of phase coordinates is now defined to be

$$\bar{f} = \frac{\int P f d\mathcal{T}}{\int P d\mathcal{T}} \quad . \quad . \quad . \quad . \quad . \quad [8]$$

the integrals now being over all of phase space. The element of volume  $d\mathcal{T}$  can be represented as

$$d\mathcal{T} = dsdn = \frac{ds dG}{|\nabla G|}$$

having the same meaning as in the previous Chapter. Therefore, with this result and the equation [7], the expression [8] goes over to

$$\begin{aligned} \bar{f} &= \frac{\int P_0 f \delta(G - G_0) ds \frac{dG}{|\nabla G|}}{\int P_0 \delta(G - G_0) ds \frac{dG}{|\nabla G|}} \\ &= \frac{\int_{G_0} f \frac{ds}{|\nabla G|}}{\int_{G_0} \frac{ds}{|\nabla G|}} \quad . \quad . \quad . \quad . \quad . \quad [9] \end{aligned}$$

Note that these integrals are surface integrals over  $G = G_0$ .

Let us now choose for  $f$ , the function  $T_r$ , given by

$$T_r = x_r \frac{\partial G}{\partial x_r} = \left\{ x_r^2 - x_r x_r^* \right\}$$

We note first that since

$$\nabla G = \sum_{\mathbf{r}} \hat{x}_{\mathbf{r}} \frac{\partial G}{\partial x_{\mathbf{r}}} = \hat{n} |\nabla G| ,$$

where  $\hat{n}$  is the unit vector along the resultant and  $\hat{x}_{\mathbf{r}}$  is the unit vector along  $x_{\mathbf{r}}$  direction, the direction cosines of  $\hat{n}$  are

$$\hat{n} \cdot \hat{x}_{\mathbf{r}} = \frac{\partial G}{\partial x_{\mathbf{r}}} / |\nabla G|$$

so that

$$\begin{aligned} x_{\mathbf{r}} \frac{\partial G}{\partial x_{\mathbf{r}}} \frac{ds}{|\nabla G|} &= x_{\mathbf{r}} (\hat{n} \cdot \hat{x}_{\mathbf{r}}) ds \\ &= (\hat{n} \cdot \vec{x}_{\mathbf{r}}) ds \end{aligned}$$

Here  $\vec{x}_{\mathbf{r}}$  denotes the vector to the point  $(0, 0, \dots, 0, x_{\mathbf{r}}, \dots, 0)$ . Then if we denote the denominator in equation [9] as  $\Delta_0$ , the ensemble average of  $T_{\mathbf{r}}$  is :

$$\bar{T}_{\mathbf{r}} = \frac{\int_{s_0} \hat{n} \cdot x_{\mathbf{r}} ds}{\Delta_0} = \frac{1}{\Delta_0} \int \text{div } x_{\mathbf{r}} d\mathcal{V}$$

$$\bar{T}_{\mathbf{r}} = \frac{Y_0}{\Delta_0} , \text{ for all } \gamma. \quad \dots \dots \dots [10]$$

Here, we have used Gauss' divergence theorem and represented the volume enclosed by  $G_0$  as  $Y_0$ . This expression [10] is a constant for all species  $\gamma$  and thus the mean  $T_{\mathbf{r}}$  for any species is the same as for any other.

Now as discussed in the previous Chapter, following equation [20], the idea of equipartition prevails here as well, as evidenced by the properties of  $\bar{T}_r$ . This opens up the possibility of defining a temperature, which then enables us to define a Gibbs canonical ensemble for our system. Before doing that, however, we show by an example that our systems admits of the ergodic hypothesis as well.

We define a function  $D_r$  as

$$D_r = \frac{\partial G}{\partial x_r} = (x_r - x_r^*) \quad . \quad . \quad . \quad [11]$$

Let us calculate the ensemble average of  $D_r$ , i.e.,

$$\begin{aligned} \bar{D}_r &= \frac{\int_{\Delta_0} \frac{\partial G}{\partial x_r} \frac{ds}{|\nabla G|}}{\Delta_0} = \frac{\int (\hat{n} \cdot \hat{x}_r) ds}{\Delta_0} \\ &= 0. \end{aligned}$$

$$\text{i.e., } \overline{(x_r - x_r^*)} = 0$$

$$\text{or } \bar{x}_r = x_r^* \quad . \quad . \quad . \quad [12]$$

This shows that the ensemble average of  $x_r$  is equal to the steady state value of  $x_r$  namely  $x_r^*$ .

Now we can show that the time average of  $N_r$  is also  $q_r$ . We have from equation [1]

$$\frac{d}{dt} (\log N_r) = \epsilon_r + \sum_s \alpha_{sr} \log N_s$$

Integrating with respect to  $t$  on both sides, between the limits 0 to  $T$ , we get

$$\frac{1}{T} \log \left\{ \frac{N_r(T)}{N_r(0)} \right\} = \epsilon_r + \sum_s \alpha_{sr} \left\{ \frac{1}{T} \int_0^T \log N_s(t) dt \right\}$$

Now since  $N$  are bounded, as  $T \rightarrow \infty$ , the left hand side vanishes, and remembering that the time average of  $\log N_s$  is nothing but

$$\overline{\log N_s} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \log N_s(t) dt,$$

we get

$$0 = \epsilon_r + \sum_s \alpha_{sr} \overline{\log N_s}$$

$$\text{or, } \sum_s \alpha_{sr} \overline{\log N_s} = -\epsilon_r$$

and through equation [2], this becomes

$$\bar{x}_s = x_s^*$$

$$\text{or } \bar{x}_r = x_r^* \quad \dots \dots \dots [13]$$

Thus, by combining equations [12] and [13], we see that the time average of  $x_r$  is precisely equal to the ensemble average of the same. Clearly, our biological system is ergodic.

## 2. THE CANONICAL ENSEMBLE

We now define a Canonical ensemble for a subsystem of our general biological system, which would consist of, say,  $\nu$  number of species, where  $1 \leq \nu \leq n$ . This subsystem has to stay in statistical equilibrium with the rest of the system. The statistical equilibrium between the parts is ensured by establishing that they all have the same 'temperature  $\theta$ '.

The density function  $P$  for such a canonical system is given by

$$P_{\nu} = e^{-G_{\nu}/\theta} \quad \dots \quad [14]$$

where, the suffix  $\nu$  indicates that we are dealing with the part containing  $\nu$  number of species. The Canonical average of any quantity  $D_r$ , where  $r$  may refer to any of the  $\nu$  species, is given by

$$\bar{D}_r = \frac{\int D_r e^{-G_{\nu}/\theta} d\mathcal{L}_{\nu}}{\int e^{-G_{\nu}/\theta} d\mathcal{L}_{\nu}} \quad \dots \quad [15]$$

Let us take

$$D_r = \frac{\partial G}{\partial x_r} = (x_r - x_r^*) \quad \dots \quad [16]$$

The canonical average of the quantity  $D_r$  is then given by



$$\begin{aligned}\bar{D}_r &= \int \frac{\partial G_r}{\partial x_r} e^{-G_r/\theta} d\tau / \int e^{-G/\theta} d\tau \\ &= \frac{-\theta \int \frac{\partial}{\partial x_r} \left\{ e^{-G_r/\theta} \right\} dx_r}{\int e^{-G_r/\theta} dx_r} \\ &= 0\end{aligned}$$

$$\text{i.e., } \bar{D}_r = (x_r - x_r^*) = 0$$

$$\text{or, } \bar{x}_r = x_r^* \quad \dots \dots \dots [17]$$

Let us now consider the ensemble average of the quantity  $D_r^2$ . We have

$$\begin{aligned}\bar{D}_r^2 &= \int \left\{ \frac{\partial G}{\partial x_r} \right\}^2 e^{-G/\theta} d\tau / \int e^{-G/\theta} d\tau \\ &= -\theta \int \left\{ \frac{\partial G_r}{\partial x_r} \right\} \frac{\partial}{\partial x_r} (e^{-G_r/\theta}) dx_r / \int e^{-G_r/\theta} dx_r \\ &= \frac{\left\{ -\theta \left[ \frac{\partial G_r}{\partial x_r} e^{-G_r/\theta} \right]_{-\infty}^{+\infty} + \theta \int e^{-G_r/\theta} \frac{\partial^2 G_r}{\partial x_r^2} dx_r \right\}}{\int e^{-G_r/\theta}} \\ &= \theta.\end{aligned}$$

The suffix  $\nu$  has been dropped here since it is unimportant: the number  $\nu$  could be anything from 1 to  $n$  so long as it contains  $r$ . This result

$$\Theta = \overline{D_r^2} = \overline{(x_r - x_r^*)^2} \quad \dots \quad [18]$$

Thus shows that the temperature  $\Theta$  is a measure of the mean square deviation of  $x_r$  from its equilibrium value  $x_r^*$ . In an exactly similar way, as in the previous Chapter, the situation of zero temperature would thus correspond to the completely quiet state of our biological system. In other words,  $\Theta$  is a measure of the level of excitation of the system from its stationary state.

Now let us consider the quantity

$$\begin{aligned} Z &= \int e^{-G/\Theta} dx_1, \dots, dx_n \\ &= \prod_r \int e^{-G_r/\Theta} dx_r \\ &= \prod_r Z_r, \quad r = 1, 2, \dots, n \quad \dots \quad [19] \end{aligned}$$

So we have

$$\begin{aligned} Z_r &= \int e^{-G_r/\Theta} dx_r \\ &= \int e^{-\left\{ \frac{1}{2} x_r^2 - x_r x_r^* \right\} / \Theta} \\ &= \sqrt{2\pi\Theta} e^{(x_r^*)^2/2\Theta} \quad \dots \quad [20] \end{aligned}$$

Note that the expression [19] is the well known Gibbs phase integral.

### 3. THE LOGNORMAL DISTRIBUTION

We are now in a position to calculate the probability  $P(x_r) dx_r$  for any species  $r$  to have its  $x_r$  in the interval between  $X_r$  and  $X_r + dX_r$ . For this, we either consider the system with  $K$  species or a one component system with species  $r$ . In the former case, after a simple integration over all the coordinates other than  $X_r$ , and in the latter case directly, we get the result

$$\begin{aligned}
 P(x_r) dx_r &= \frac{e^{-G_r/\Theta} dx_r}{Z_r} \\
 &= \frac{e^{-\frac{1}{2\Theta} \left\{ \frac{x_r^2}{2} - x_r x_r^* \right\}} dx_r}{\sqrt{2\pi\Theta} e^{(x_r^*)/2\Theta}} \\
 &= \frac{1}{\sqrt{2\pi\Theta}} e^{-\frac{1}{2\Theta} [x_r - x_r^*]^2} dx_r \dots [21]
 \end{aligned}$$

where, of course

$$\Theta = \frac{1}{(x_r - x_r^*)^2} \dots \dots \dots [22]$$

By transforming back to  $N_r$ , through the relation [3], this may also be written as

$$P(N_r) dN_r = \frac{1}{N_r \sqrt{2\pi\Theta}} e^{-\frac{1}{2\Theta} [\log N_r - \log N_r^*]^2} dN_r \dots [23]$$

where

$$\Theta = \overline{(\log N_r - \log N_r^*)^2} \dots [24]$$

The expression in [23] is the standard lognormal distribution.

#### 4. COMPARISON WITH THE RESULT OF PRESTON

Let us now consider the approximation in which the equilibrium value of the number of individuals for any species within the genera is the same. In that case the function  $P(x_r)$  multiplied by  $N$ , which is the total number of species in the system, has precisely the same meaning as the function  $n(R)$  or  $n(\log^1)$  of Preston. In other words, we have the result

$$n(\log^1) = \frac{N}{\sqrt{2\pi\Theta}} e^{-\frac{1}{2\Theta} [x - x^*]^2} \dots [25]$$

where  $\Theta$  has the same meaning as  $R^2$  or  $\sigma^2$  in Preston's notation, and of course

$$x = \log^1,$$

$i$  here is the number of individuals referred to by  $N_r$  so far in this Chapter. Clearly, the suffix  $r$  has been dropped because the distribution is now the same for any  $r$ . We have thus established Preston's result for the frequency distribution of species within a genera. [Note

that the logarithms in [25] are to the base  $e$ , whereas in the expression of Preston given in equation [7] of Chapter III, the logarithms are to the base 2. The latter can, however, be converted to the base  $e$  without any change in its form, except for an overall scale factor of  $\log_2 e$  ].

REFERENCES

1. Corbet, A.S., *Journal of Animal Ecology* 12, 42-58 (1943).  
See also Corbet, A.S., *Proc. R. Ent. Soc. Lond.*, (A), 16, 101-16 (1942).
2. Williams, C.B., *Journal of Animal Ecology* 12, 42-58 (1943).  
See also Williams, C.B., *Patterns in the Balance of Nature*. Academic, New York (1964).
3. For Saunders' observations, see reference No. 8.
4. For Preston and Norris's data, see reference No. 8.
5. For Dirk's data, see reference No. 8.
6. For Seaman's data, see reference No. 8.
7. Fisher, R.A., Corbet, A.S. and Williams, C.B., *Journal of Animal Ecology*, 12, 42-58 (1943).
8. Preston, F.W., *Ecology*, 29, 254-283 (1948).
9. Kendall, D.G., *Biometrika*, 35, 6-15 (1948).
10. Kerner, E.H., *Bulletin of Mathematical Biophysics*, 19, 121-46 (1957).
11. Bhat, N. and Pande, L.K., *Journal of Theoretical Biology*, 83, 321-344 (1980).