A MODEL FOR THE DISTRIBUTION OF INDIVIDUALS BY SPECIES IN AN ENVIRONMENT

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

John W. McCloskey

1965

[redacted]

THESIS

This is to certify that the

thesis entitled

A MODEL FOR THE DISTRIBUTION
OF INDIVIDUALS BY SPECIES
IN AN ENVIRONMENT

presented by

John W. McCloskey

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

Herman Rubin
Major professor

Date May 17, 1965

O-169

ABSTRACT

A MODEL FOR THE DISTRIBUTION
OF INDIVIDUALS BY SPECIES
IN AN ENVIRONMENT

by John W. McCloskey

The problem considered in this thesis is that of developing a model for biological environments so that, for samples of individuals obtained from the environment, the number of species and the number of individuals in the respective species can be predicted. It is assumed that the number of individuals in the environment, as well as the number of species, is countably infinite so that only in environments where these quantities are very large will the model be realistic.

In Chapter 1 the model is developed and in Chapter 2 a procedure developed to obtain maximum likelihood estimates of the parameters of the model using a sample of data already gathered from the environment. Since there are three parameters in the model the estimates are obtained from the

simultaneous solution of three equations which is accomplished by means of an iterative Newton procedure.

As a means of studying the behavior of the model a simulation procedure was developed in Chapter 4 which would choose a sample from the model for a given set of parameters. This procedure uses random variables having Binomial, Poisson, Hypergeometric, Truncated Poisson and Exponential distributions. Methods were thus developed in Chapter 3 to produce random variables with these specified distributions rapidly and with as few input random variables as possible. The fundamental technique used in obtaining these random variables is the acceptance-rejection technique introduced by von Neumann.

Chapter 5 and Chapter 6 are devoted to the analysis of data that was taken from actual biological environments. The analysis is accomplished through procedures developed in the previous chapters and the Control Data 3600 computer used for the actual calculations. Several FORTRAN 60 programs were used for these calculations which are tabulated in the appendix.

A MODEL FOR THE DISTRIBUTION

OF INDIVIDUALS BY SPECIES

IN AN ENVIRONMENT

By

John W. McCloskey

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics

1965

ACKNOWLEDGMENTS

I would like to thank my academic advisor, Professor Herman Rubin, for his many comments and suggestions in the course of the research. His willingness to discuss the problem and to exchange ideas in the early stages of the research was especially helpful. Also, the ideas expressed by Professor Philip Clark concerning the presentation of the simulated data were very helpful and greatly appreciated.

I am also grateful to the Department of Statistics, Michigan State University, and the Office of Naval Research for their financial support during the writing of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Section 1: General Discussion of the Model

Let C be the collection of all the individuals of a certain type, for example butterflies, present in an environment. Consider the partition of the individuals into species designated by $\{s_1, s_2, \ldots\}$ where the species are arbitrarily named $s_1, s_2, \ldots$ and suppose the number of species present in the environment is countably infinite. This assumption is made because in the environments being considered the number of species is very large and in sampling from the environment there is assumed to be a strictly positive probability of finding a new species regardless of the number of species that have already been found. Also define a probability p_i for each species s_i such that $\sum_{i=1}^{\infty} p_i = 1$.

Consider now the task of choosing a sample of N individuals independently from the environment. Let these individuals be designated by $I_1, I_2, \ldots, I_N$. The individuals are chosen according to the restriction

$$P[I_i \text{ is from the species } s_j] = p_j \qquad \text{for } i = 1,2,\ldots,N \quad j = 1,2,\ldots$$

After the individuals are chosen from the environment the sample will contain say s species for which there are n_1 species with one individual, n_2 species with two individuals and in general n_i species with i individuals subject to the conditions

$$\sum_{i=1}^{N} n_i = s \quad \text{and} \quad \sum_{i=1}^{N} i\, n_i = N.$$

The object of this report is to develop a model for natural environments so that the distribution of the numbers $s, n_1, n_2, \ldots, n_N$ can be predicted.

Consider therefore the generalization of the probabilities p_i where for each species s_i in the environment it is assumed there is an "intensity" x_i proportional to p_i. Let $z = \sum_{i=1}^{\infty} x_i$ where z is defined to be the total intensity of the environment. Define an intensity function f to be a non-negative integrable function on $[0,\infty)$ with the property that (i) for any $\epsilon > 0$, $\int_0^{\epsilon} f(x)dx = +\infty$ and $\int_{\epsilon}^{\infty} f(x)dx < +\infty$ and (ii) $\int_0^{\infty} xf(x)dx < +\infty$.

The model can now be stated as follows: Given an intensity function f for an environment, for any interval $[a,b)$ with $0 < a < b \leq +\infty$ the number of species present with intensity x_i in the interval $a \leq x_i < b$ has a Poisson distribution with mean $\int_a^b f(x)dx$ and for disjoint intervals the number of species with intensities in the respective intervals have independent Poisson distributions. Condition (i) on f is made so that the expected number of species will be infinite and (ii) is made so that the total intensity will be almost surely finite. Let U_i be a random variable representing the number of individuals observed from species s_i for $i = 1,2,\ldots$. Suppose $U_1,U_2,\ldots$ to be independent Poisson random variables with means $k_s x_i$, where k_s is a positive constant and x_i the intensity of the respective species. Define a sample to be an observation of the random vector $U = (U_1,U_2,\ldots)$ and define Y_m = (number of $U_i = m$) for $m = 1,2,\ldots$.

The development which follows in this section is an attempt to give motivation for the actual development of the model in the next section. Thus, let $X = (X_1,X_2,\ldots)$ be a set of intensities obtained from the process and define $Z = \sum_{i=1}^{\infty} X_i$ and the species with intensity X_i will be designated species s_i.

Then

$P[I_1 \text{ is from species } s_j] = \frac{X_j}{Z}$ for $j = 1,2,\ldots$. Let S_1^* be the species of individual I_1 and let V_1 be the intensity of this species. Choose a second individual I_2 randomly from the environment and examine its species. If it is different from S_1^*, let S_2^* be its species and V_2 the intensity of this species. If however I_2 is from the same species as I_1 continue selecting individuals independently from the environment until one is found which has a different species than I_1 and let the species of this individual be S_2^*. Consider now the two random variables

$$W_1 = \frac{V_1}{Z} \qquad \text{and} \qquad W_2 = \frac{V_2}{Z-V_1} .$$

Theorem 1: Suppose that W_1 and W_2 are independent and identically distributed according to a distribution H on [0,1]. If $0 < E(W_1) < 1$, then define λ such that $E(W_1) = \frac{1}{\lambda+1}$. It then follows that $d\,H(w) = \lambda(1-w)^{\lambda-1}dw$.

Proof: Let Y_{I_i} for $i = 1,2$, be the proportion of individuals in the environment from the same species as individual I_1. The individuals I_1 and I_2 are chosen independently from the same environment so Y_{I_1} and Y_{I_2} are independent and identically distributed. Y_{I_i} is defined as follows

$$Y_{I_1} = W_1$$

and

$$Y_{I_2} = \begin{cases} W_1 \text{ with probability } W_1 \\ (1-W_1)W_2 \text{ with probability } (1-W_1). \end{cases}$$

Let the r^{th} moment of H be μ_r. Then for $r > 0$

$$\mu_r = E(W_1{}^r) = E(Y_{I_1}^r) = E(Y_{I_2}^r)$$

$$= E[W_1{}^{r+1} + W_2{}^r (1-W_1)^{r+1}]$$

$$= \mu_{r+1} + \mu_r E(1-W_1)^{r+1}$$

$$= \mu_{r+1} + \mu_r \left[\sum_{k=0}^{r+1} (-1)^k \binom{r+1}{k} \mu_k\right]$$

Solving for μ_{r+1}

$$\mu_{r+1} = \frac{\mu_r \left[1 - \sum_{k=0}^{r} (-1)^k \binom{r+1}{k} \mu_k\right]}{1 + (-1)^{r+1} \mu_r}$$

From this equation μ_{r+1} is determined by $\mu_0, \mu_1, \ldots, \mu_r$ unless r is even and $\mu_r = 1$. If however $\mu_r = 1$ for $r > 0$ the distribution is concentrated at one and all $\mu_k = 1$. This distribution with $\mu_k = 1$ for all k indicates that all individuals are from the same species which violates the assumption that the environment contain an infinite number of species.

In order to determine the moments μ_r an equality must first be established. Thus

$$\int_0^1 (1-x)^{r+1} (1-x)^{\lambda-1} dx = \int_0^1 \sum_{k=0}^{r+1} (-1)^k \binom{r+1}{k} x^k (1-x)^{\lambda-1} dx$$

$$= \sum_{k=0}^{r+1} (-1)^k \binom{r+1}{k} \int_0^1 x^k (1-x)^{\lambda-1} dx$$

$$= \sum_{k=0}^{r+1} (-1)^k \binom{r+1}{k} \frac{\Gamma(k+1)\Gamma(\lambda)}{\Gamma(k+1+\lambda)} = \sum_{k=0}^{r+1} (-1)^k \binom{r+1}{k} k! \frac{\Gamma(\lambda)}{\Gamma(k+1+\lambda)}$$

Thus

$$\sum_{k=0}^{r} (-1)^k \binom{r+1}{k} k! \frac{\Gamma(\lambda)}{\Gamma(k+1+\lambda)} = \int_0^1 (1-x)^{r+\lambda} dx - (-1)^{r+1}(r+1)! \frac{\Gamma(\lambda)}{\Gamma(r+2+\lambda)}$$

$$= \frac{1}{r+\lambda+1} - (-1)^{r+1} (r+1)! \frac{\Gamma(\lambda)}{\Gamma(r+2+\lambda)}$$

It is to be shown now with the use of the above equation that $\mu_k = \frac{k!\Gamma(\lambda+1)}{\Gamma(k+1+\lambda)}$ by induction. Obviously $\mu_1 = \frac{1}{\lambda+1}$ and assume

$\mu_k = \frac{k!\Gamma(\lambda+1)}{\Gamma(k+1+\lambda)}$ for $k = 0,1,2,\ldots r$.

From the recursion formula for μ_{r+1}

$$\mu_{r+1} = \frac{r!\Gamma(\lambda+1)}{\Gamma(k+1+\lambda)} \left[\frac{1 - \sum_{k=0}^{r} (-1)^k \binom{r+1}{k} \frac{k!\Gamma(\lambda+1)}{\Gamma(k+1+\lambda)}}{1 + (-1)^{r+1} r! \frac{\Gamma(\lambda+1)}{\Gamma(r+1+\lambda)}} \right]$$

$$= \frac{r!\,\Gamma(\lambda+1)}{\Gamma(k+1+\lambda)} \left[\frac{1 - \lambda \frac{1}{r+1+\lambda} - (-1)^{r+1} (r+1)! \frac{\Gamma(\lambda)}{\Gamma(r+2+\lambda)}}{1 + (-1)^{r+1} r! \frac{\Gamma(\lambda+1)}{\Gamma(r+1+\lambda)}} \right]$$

$$= \frac{r!\,\Gamma(\lambda+1)}{\Gamma(k+1+\lambda)} \left[\frac{\frac{r+1}{r+1+\lambda} + (-1)^{r+1} \frac{(r+1)}{r+1+\lambda} r! \frac{\Gamma(\lambda+1)}{\Gamma(r+1+\lambda)}}{1 + (-1)^{r+1} r! \frac{\Gamma(\lambda+1)}{\Gamma(r+1+\lambda)}} \right]$$

$$= \frac{(r+1)!\,\Gamma(\lambda+1)}{\Gamma(r+2+\lambda)}$$

Consider the r^{th} moment of the distribution $\lambda(1-x)^{\lambda-1}$ for $0 \leq x \leq 1$, 0 otherwise

$$\int_0^1 x^r \lambda(1-x)^{\lambda-1} dx = \frac{\lambda\, \Gamma(r+1)\, \Gamma(\lambda)}{\Gamma(r+1+\lambda)} = \frac{r!\, \Gamma(\lambda+1)}{\Gamma(r+1+\lambda)}.$$

This distribution has the desired moments and since its moment generating function exists in a neighborhood of zero

$$dH(w) = \lambda(1-w)^{\lambda-1}\, dw.$$

Chapter 1

Section 2: Development of the Model

Let f be an intensity function and let $X = (X_1, X_2, X_3, \dots)$ be a set of intensities obtained from the process described in the previous section using f as the intensity function. Let $Z = \sum_{i=1}^{\infty} X_i$ and let Z have density g. Define V_1 to be a random variable such that $P(V_1 = X_i | X) = \frac{X_i}{Z}$ for i = 1,2,... and define $Y = Z - V_1$.

Lemma 1: If f is a continuous intensity function the joint density h of V_1 and Y can be expressed in the form

$$h(v_1, y) = \frac{v_1 f(v_1)}{v_1 + y} g(y) .$$

The proof of this result was obtained by Professor Herman Rubin and is to be published in a paper by him.

Make the substitution $z = v_1 + y$ so that $h_{V_1,Z}(v_1, z) = \frac{v_1 f(v_1)}{z} g(z - v_1)$. Now integrating with respect to v_1 to obtain the density of the total intensity

$$g(z) = \int_0^z h_{V_1,Z}(v_1, z) dv_1 = \int_0^z \frac{v_1 f(v_1)}{z} g(z - v_1) dv_1 .$$

Define $W = \frac{V_1}{Z}$. Then

$$h_{W,Z}(w,z) = \frac{wzf(wz)g(z-zw)z}{z} = wzf(wz)g(z(1-w)).$$

Theorem 2: If $h_{W,Z}(w,z) = wzf(wz)g(z(1-w))$ for $0 \leq w \leq 1$ and $0 < z < \infty$ and if $h_{W|z}(w) \overset{z}{\equiv} \varphi(w)$ and assuming f and g to be twice differentiable then $f(x) = c\, x^{-1} e^{kx}$ for $0 < x < \infty$ and $g(z) = c' z^{H'} e^{kz}$ for $0 < z < \infty$.

Proof: $h_{W|z}(w) = \frac{h_{W,Z}(w,z)}{g(z)} = \frac{wz\ f(wz)\ g(z(1-w))}{g(z)} = \varphi(w)$

taking logarithms

$$\log w + \log z + \log f(wz) + \log g(z(1-w))$$

$$= \log \varphi(z) + \log g(z)$$

Let $\psi_1(wz) = \log f(wz)$

and $\psi_2(z(1-w)) = \log g(z(1-w))$

thus

$$\log w + \log z + \psi_1(wz) + \psi_2(z(1-w)) = \log \varphi(w) + \log g(z)$$

taking derivative with respect to w and then with respect to z

$$\frac{1}{w} + z\,\psi_1'(wz) - z\,\psi_2'(z(1-w)) = \frac{\varphi'(w)}{\varphi(w)}$$

$$\psi_1'(wz) + wz\,\psi_1''(wz) - \psi_2'(z(1-w)) - z(1-w)\,\psi_2''(z(1-w)) = 0.$$

Thus

$$wz\,\psi_1''(wz) + \psi_1'(wz) = z(1-w)\,\psi_2''(z(1-w)) + \psi_2'(z(1-w))$$

Since the above equation is valid for all values of z and w the following must be true

$$wz\,\psi_1''(wz) + \psi_1'(wz) = k$$

and $z(1-w)\,\psi_2''(z(1-w)) + \psi_2'(z(1-w)) = k.$

Solving then these two differential equations

$$u\,\psi_1''(u) + \psi_1'(u) = k$$

$$u\,\psi_1'(u) = ku + H$$

$$\psi_1'(u) = k + \frac{H}{u}$$

$$\psi_1(u) = ku + H \log u + M$$

$$f(u) = e^{\psi_1(u)} = c\,u^H\,e^{ku}$$

Similarly

$$g(v) = e^{\psi_2(v)} = c'\,v^{H'}\,e^{kv}$$

Finding now the particular solution

$$\varphi(w) = h_{W|z} = \frac{wz\ f(wz)\ g(z(1-w))}{g(z)}$$

$$= wz\ c\ w^{H}\ z^{H}\ e^{kwz}\ \frac{c'\, z^{H'}\ (1-w)^{H'}\ e^{kz(1-w)}}{c' z^{H'}\ e^{kz}}$$

$$= c\ w^{H+1}\ z^{H+1} (1-w)^{H'}$$

which implies that $H = -1$ yielding the final result $f(u)=c\ u^{-1}e^{ku}$.

From the above analysis and in an effort to make the model as general as possible the form of the function f was decided to be $f(x) = \frac{A\ e^{-cx}}{x^{\alpha}}$.

Obviously $A > 0$ and due to the restrictions of the model $c \geq 0$ since $\int_N^{\infty} f(x)dx \to 0$ as $N \to \infty$ because the total intensity of the large species is almost surely finite. Also $\alpha \geq 1$ because if $\alpha < 1$ then $\int_0^{\epsilon} f(x)dx = \int_0^{\epsilon} \frac{Ae^{-cx}}{x^{\alpha}}\, dx \leq \int_0^{\epsilon} \frac{A}{x^{\alpha}}\, dx = \frac{A\epsilon^{1-\alpha}}{1-\alpha} < \infty$ controdicting the restriction that the expected number of small species present be infinite.

From the development in Chapter 2 it can easily be observed that the transformation $x \to \lambda x$, $c \to c/\lambda$, $k_s \to k_s/\lambda$, $A \to A/\lambda^{1-\alpha}$ preserves the model so that only α, $k_s/(k_s+c)$ and $A/(k_s+c)^{1-\alpha}$ are identifiable. For this reason only the cases $c = 0$ and $c = 1$ need be considered. The general form of the function f was taken to be $f(x) = \frac{Ae^{-x}}{x^{\alpha}}$ for the work which immediately follows while in Chapter 6 a generalized form of the case $c = 0$ is considered.

Chapter 1

Section 3: A Sepcial Case of the Model

Consider now a special case of the model developed in the previous section where $f(x) = \frac{Ae^{-x}}{x}$. Knowing g(z) has the form $g(z) = c'z^{H'}e^{-z}$ and using the previously established equation for g(z),

$$g(z) = \int_0^z h_{V_1,Z}(v_1,z)dv_1 = \int_0^z \frac{v_1 f(v_1)}{z} g(z-v_1)dv_1 \quad .$$

$$= \int_0^z \frac{v_1 Ae^{-v_1}}{v_1 z} \; c'(z-v_1)^{H'} \; e^{-z+v_1} dv_1$$

$$= \frac{Ac'e^{-z}}{z} \int_0^z (z-v_1)^{H'} dv_1 = \frac{Ac'e^{-z}z^{H'+1}}{z(H'+1)} .$$

This equation implies $H' = A - 1$ and since $\int_0^\infty c' z^{A-1} e^{-z} dz = \Gamma(A)$, then $c' = \frac{1}{\Gamma(A)}$.

Therefore $g(z) = \frac{1}{\Gamma(A)} z^{A-1} e^{-z}$.

For $j = 1,2,\ldots$ define V_j to be a random variable such that

$$P(V_j = X_i \mid X) = \frac{X_i}{Z - \sum_{i=1}^{j-1} V_i} \quad \text{for all } i \text{ except}$$

those i's for which $X_i = V_k$ for $k = 1,2,\ldots j-1$.

Let

$$W_i = \frac{V_i}{Z - \sum_{j=1}^{i-1} V_j} .$$

By repeated application of the formula $g(z) = \int_0^z \frac{xf(x)}{z} g(z-x)dx$ which was previously established

$$g(z) = \int_0^z \frac{v_1 f(v_1)}{z} \int_0^{z-v_1} \frac{v_2 f(v_2)}{(z-v_1)} \cdots \int_0^{z-\sum_{j=1}^{i-1} v_j} \frac{v_i f(v_i)}{(z-\sum_{j=1}^{i-1} v_j)} g(z-\sum_{j=1}^{i} v_j)dv_i \ldots dv_1$$

$$= \int_0^z \int_0^{z-v_1} \cdots \int_0^{z-\sum_{j=1}^{i-1} v_j} \frac{v_1 f(v_1)}{z} \frac{v_2 f(v_2)}{(z-v_1)} \cdots \frac{v_i f(v_i)}{(z-\sum_{j=1}^{i-1} v_j)} g(z-\sum_{j=1}^{i} v_j) dv_i \ldots dv_1$$

so the joint density for $V_1, V_2, \ldots, V_i, Z$ where $v_0 = 0$ becomes

$$h_{V_1,V_2,\ldots,V_i,Z}(v_1,v_2,\ldots,v_i,z) = \left[\prod_{j=1}^{i} \frac{v_j f(v_j)}{\left(z-\sum_{k=0}^{j-1} v_k\right)} \right] g(z-\sum_{j=1}^{i} v_j).$$

Theorem 3: In an environment where $f(x) = \frac{Ae^{-x}}{x}$ and $g(z) = \frac{1}{\Gamma(A)} z^{A-1} e^{-z}$ and where V_i, W_i, Z and the joint density $h_{V_1,V_2,\ldots,V_i,Z}$ are defined as above, W_i is distributed according to the distribution

$$h^*(w_i) = A(1-w_i)^{A-1} \qquad \text{for} \quad 0 \le w_i \le 1.$$

Proof: For $i \ge 3$ the joint density

$$h_{V_1,V_2,\ldots,V_i,Z}(v_1,v_2,\ldots,v_i,z) = \left[\prod_{j=1}^{i} \frac{v_j f(v_j)}{\left(z-\sum_{k=0}^{j-1} v_k\right)} \right] g(z-\sum_{j=1}^{i} v_j)$$

$$= \frac{A^i}{\Gamma(A)} \left[\prod_{j=1}^{i} \frac{1}{\left(z-\sum_{k=0}^{j-1} v_k\right)} \right] (z - \sum_{j=1}^{i} v_j)^{A-1} e^{-z}$$

$$= \frac{A^i}{\Gamma(A)} \left[\prod_{j=1}^{i-1} \frac{1}{\left(z-\sum_{k=0}^{j-1} v_k\right)} \right] (z-\sum_{j=1}^{i-1} v_j)^{A-2} \left(1 - \frac{v_i}{z-\sum_{j=1}^{i-1} v_j}\right)^{A-1} e^{-z}.$$

Make the substitution $W_i = \dfrac{V_i}{Z - \sum_{j=1}^{i-1} V_j}$ so that

$$h_{V_1,V_2,\ldots,V_{i-1}W_i,Z}(v_1,\ldots,v_{i-1}w_i,z) =$$

$$\frac{A^i}{\Gamma(A)}\left[\prod_{j=1}^{i-1} \frac{1}{\left(z-\sum_{k=0}^{j-1} v_k\right)}\right]\left(z - \sum_{j=1}^{i-1} v_j\right)^{A-1} (1-w_i)^{A-1} e^{-z}.$$

Integrating this density then

$$h_{W_i,Z}(w_i,z) = \int_0^{z} \cdots \int_0^{z-\sum_{j=1}^{i-2} v_j} h_{V_1,V_2,\ldots,V_{i-1},W_i,Z}(v_1,\ldots,v_{i-1},w_i,z)\,dv_{i-1}\ldots dv_1$$

$$= \frac{A}{\Gamma(A)} z^{A-1} (1-w_i)^{A-1} e^{-z}.$$

Integrating now with respect to z, $h^*(w_i) = \int_0^\infty h_{W_i,Z}(w_i,z)\,dz$

$= \int_0^\infty \frac{A}{\Gamma(A)} (1-w_i)^{A-1} z^{A-1} e^{-z}\,dz = A(1-w_i)^{A-1}$. For $i = 1,2$, the same procedure is followed with simplification in the integration.

Chapter 2

Section 1: Maximum Likelihood Estimates of Parameters

The general form of the intensity function has been established to be $f(x) = \frac{A e^{-x}}{x^{\alpha}}$ where A, α are parameters of the function. In any sample that is taken from the model the number of individuals in each species is assumed to be Poisson with mean proportional to the intensity of the species; that is the number of individuals in the sample from the i^{th} species is Poisson with mean $k_s x_i$ where k_s is defined to be the intensity of the sample. This parameter k_s is also to be estimated.

Suppose now that data is available from this model and it is desired to estimate the above parameter. Let y_m be the number of species with m individuals in the sample, I the number of individuals and s the number of species. The following trivial equations are to hold $\sum_{m=1}^{\infty} y_m = s$ and $\sum_{m=1}^{\infty} m\, y_m = I$.

In accordance with the above notation the probability that there will be m individuals in the sample from a species with intensity x is $\frac{(k_s x)^m}{m!} e^{-k_s x}$ and the expected number of species in the sample with m individuals is

$$\int_0^{\infty} \frac{(k_s x)^m}{m!} e^{-k_s x} f(x)\,dx = \int_0^{\infty} A x^{-\alpha} e^{-x} \frac{(k_s x)^m}{m!} e^{-k_s x}\,dx$$

$$= \frac{A k_s^{\,m}}{m!} \frac{\Gamma(m-\alpha+1)}{(k_s+1)^{m-\alpha+1}} = \frac{A}{(k_s+)^{1-\alpha}} \left(\frac{k_s}{k_s+1}\right)^m \frac{\Gamma(m-\alpha+1)}{m!} = B\eta^m \frac{\Gamma(m-\alpha+1)}{m!}$$

by making the substitution $\eta = \left(\frac{k_s}{k_s+1}\right)$ and $B = \frac{A}{(k_s+1)^{1-\alpha}}$.

Since the total number of species present in the sample has a Poisson distribution, the y_m are independent and have a Poisson distribution with mean $B\,\eta^m \frac{\Gamma(m-\alpha+1)}{m!}$.

The density thus becomes

$$f(y_1,y_2,y_3,\ldots;B,\eta,\alpha) = \prod_{m=1}^{\infty} e^{-\lambda_m} \frac{\lambda_m^{y_m}}{y_m!}$$ where y_m is as previously defined and $\lambda_m = B\eta^m \frac{\Gamma(m-\alpha+1)}{m!}$, the expected number of species in the sample with m individuals.

The logarithm of the density as a function of the three parameters ignoring constants becomes

$$L(B,\alpha,\eta) = \sum_{m=1}^{\infty} - B\eta^m \frac{\Gamma(m-\alpha+1)}{m!} + \sum_{m=1}^{\infty} y_m[\log B + m \log \eta + \log \Gamma(m-\alpha+1) - \log m!].$$

Now simplifying the first term

$$\sum_{m=1}^{\infty} - B\eta^m \frac{\Gamma(m-\alpha+1)}{m!} = \sum_{m=1}^{\infty} - \frac{B\eta^m}{m!} \int_0^{\infty} x^{m-\alpha} e^{-x} dx$$

$$= \int_0^{\infty} \sum_{m=1}^{\infty} - \frac{B\eta^m}{m!} x^m x^{-\alpha} e^{-x} dx = -B \int_0^{\infty} x^{-\alpha} e^{-x} (e^{\eta x} - 1) dx$$

From Bierens DeHaan [1] table #90 equation #6

$$\int_0^{\infty} (e^{-qx} - e^{-rx}) \frac{dx}{x^{p+1}} = \frac{1}{p} \Gamma(1-p) (r^p - q^p) \quad \text{for } p < 1.$$

Let $p = \alpha - 1$. Then $-B \int_0^{\infty} (e^{-(1-\eta)x} - e^{-x}) x^{-\alpha} dx =$

$$-B \frac{1}{\alpha-1} \Gamma(2-\alpha)(1-(1-\eta)^{\alpha-1}) = -B\Gamma(1-\alpha) \left[(1-\eta)^{\alpha-1} - 1 \right].$$

Using the above and simplifying the second term, the likelihood function thus becomes $L(B,\alpha,\eta) =$

$$-B\Gamma(1-\alpha)\left[(1-\eta)^{\alpha-1} - 1 \right] + s \log B + I \log \eta + \sum_{m=1}^{\infty} y_m \log\Gamma(m-\alpha+1) - \sum_{m=1}^{\infty} y_m \log m!$$

It was found that in taking the derivative of the above function with respect to the parameter η the resulting equation was very unstable for η near one and α near one. To eleviate this difficulty the substitution $(1-\eta) = e^{-q}$ was made. The likelihood function $L(B,\alpha,q)$ thus becomes

$$L(B,\alpha,q) = -B\Gamma(1-\alpha)\left[e^{-q(\alpha-1)}-1\right] + s \log B + I\log(1-e^{-q})$$
$$+ \sum_{m=1}^{\infty} y_m \log \Gamma(m-\alpha+1) - \sum_{m=1}^{\infty} y_m \log m!.$$

Taking the derivatives with respect to these parameters

$$L_B = \frac{\partial L}{\partial B} = -\Gamma(1-\alpha)\left[e^{-q(\alpha-1)}-1\right] + \frac{s}{B}$$

$$L_q = \frac{\partial L}{\partial q} = -B\,\Gamma(1-\alpha)\,(1-\alpha)\,e^{-q(\alpha-1)} + I\frac{e^{-q}}{1-e^{-q}}$$

$$= -B\,\Gamma(2-\alpha)\,e^{-q(\alpha-1)} + I\frac{1}{e^{q}-1}$$

and using the notation

$$\psi(x) = \frac{\partial}{\partial x}\log\Gamma(x) = \frac{1}{\Gamma(x)}\frac{\partial}{\partial x}\Gamma(x)$$

so that

$$\frac{\partial}{\partial x}\Gamma(x) = \Gamma(x)\,\psi(x)$$

then

$$L_\alpha = \frac{\partial L}{\partial \alpha} = B\Gamma(1-\alpha)\,\psi(1-\alpha)\,[e^{-q(\alpha-1)}-1]$$

$$+ qB\Gamma(1-\alpha)\,e^{-q(\alpha-1)} - \sum_{m=1}^{\infty} y_m\,\psi(m-\alpha+1)$$

In finding a solution for the equations $L_\alpha = L_q = L_B = 0$ a Newton approximation in three variables was first attempted but abandoned since the matrix involved in using this method is almost singular causing instability in the procedure.

Therefore the following modified Newton method in two variables was used:

1. Initial estimates $\hat{\alpha}_1$ and $\hat{q}_1$ are given
2. Solve equation $L_B = 0$ for B to get initial estimate $\hat{B}_1$
3. Step two makes $L_B(\hat{B}_1, \hat{q}_1, \hat{\alpha}_1) = 0$ so that

$$\begin{pmatrix} 0 \\ -Lq(\hat{B}_1, \hat{q}_1, \hat{\alpha}_1) \end{pmatrix} = \begin{pmatrix} L_{BB} & L_{Bq} \\ L_{qB} & L_{qq} \end{pmatrix} \begin{pmatrix} \Delta B \\ \Delta q \end{pmatrix}$$

can be solved for Δq as follows

$$\Delta q = \frac{-L_{BB}L_q}{L_{BB}L_{qq} - L_{qB}L_{Bq}}$$

4. $\hat{q}_2 = \hat{q}_1 + \Delta q$
5. Solve $L_B = 0$ using estimates $\hat{\alpha}_1$ and $\hat{q}_2$ to obtain $\hat{B}_2$
6. As in step 3 find $\Delta\alpha$ by the equation

$$\Delta\alpha = \frac{-L_{BB}L_\alpha}{L_{BB}L_{\alpha\alpha} - L_{\alpha B}L_{B\alpha}}$$

7. $\hat{\alpha}_2 = \hat{\alpha}_1 + \Delta\alpha$
8. Continue iterating until desired accuracy is reached. This procedure gives likelihood estimates $\hat{\alpha}$, $\hat{B}$ and $\hat{q}$ from which can be calculated the other two parameters

$$\hat{k}_s = \frac{\eta}{1-\eta} = \frac{1-e^{-\hat{q}}}{e^{-\hat{q}}} = e^{\hat{q}} - 1$$

and

$$\hat{A} = \hat{B}(\hat{k}_s + 1)^{1-\hat{\alpha}} .$$

The second derivatives of the likelihood function necessary for the above method are as follows:

$$L_{BB} = \frac{\partial^2 L}{\partial B \partial B} = \frac{-s}{B^2}$$

$$L_{Bq} = L_{qB} = -\Gamma(2-\alpha)\ e^{-q(\alpha-1)}$$

$$L_{B\alpha} = L_{\alpha B} = -\Gamma(1-\alpha)\ \psi(1-\alpha)\ [1-e^{-q(\alpha-1)}] + q\Gamma(1-\alpha)\ e^{-q(\alpha-1)}$$

$$L_{qq} = B(\alpha-1)\ \Gamma(2-\alpha)\ e^{-q(\alpha-1)} - I \frac{e^q}{(e^q-1)^2}$$

$$L_{\alpha\alpha} = B\Gamma(1-\alpha)\ \psi^2(1-\alpha)\ [1-e^{-q(\alpha-1)}]$$

$$+ B\Gamma(1-\alpha)\ \psi'(1-\alpha)\ [1-e^{-q(\alpha-1)}]$$

$$-2\ qB\Gamma(1-\alpha)\ \psi(1-\alpha)\ e^{-q(\alpha-1)} - q^2 B\Gamma(1-\alpha)\ e^{-q(\alpha-1)}$$

$$+ \sum_{m=1}^{\infty} y_m\ \psi'(m-\alpha+1)$$

For the calculation of $\psi(x)$ and $\psi'(x)$ Stirling's asymptotic series is used for $\log \Gamma(x+1)$. Thus

$$\log \Gamma(x+1) = (x + \tfrac{1}{2}) \log x - x + \tfrac{1}{2} \log 2\pi$$

$$+ \frac{1}{12x} - \frac{1}{360x^3} + \frac{1}{1260x^5} - \frac{1}{1680x^7} + \ldots$$

$$\psi(x+1)=\frac{\partial}{\partial x}\log\Gamma(x+1)=\log x+\frac{1}{2x}-\frac{1}{12x^2}+\frac{1}{120x^4}-\frac{1}{252x^6}+\frac{1}{240x^8}\mp\ldots.$$

$$\psi'(x+1)=\frac{1}{x}-\frac{1}{2x^2}+\frac{1}{6x^3}-\frac{1}{30x^5}+\frac{1}{42x^7}-\frac{1}{30x^9}\pm\ldots.$$

For $x \geq 10$, $\psi(x)$ and $\psi'(x)$ are calculated from the above equations. However for small x the recursion formula

$$\Gamma(x+1) = x\Gamma(x) \text{ is used.}$$

$$\log\Gamma(x+1) = \log x + \log\Gamma(x)$$

Differentiating both sides

$$\psi(x+1) = \frac{1}{x} + \psi(x)$$

and $\psi'(x+1) = -\frac{1}{x^2} + \psi'(x)$

In the calculation of the Newton process it is often necessary to evaluate the expression $\Gamma(1-\alpha)\,[e^{-q(\alpha-1)}-1]$. It is often the case that α is near one which requires that this expression be evaluated with care to avoid the loss of several significant digits. For this reason make the following substitution:

$$\Gamma(1-\alpha)\,[e^{-q(\alpha-1)}-1] = \Gamma(2-\alpha)\,\frac{1-e^{-q(\alpha-1)}}{\alpha-1}$$

$$= \Gamma(2-\alpha)\,e^{-\frac{qz}{2}}\,\frac{\sinh\frac{qz}{2}}{\frac{qz}{2}}\cdot q \quad \text{where } z = \alpha - 1.$$

Now let

$$h = \frac{\tanh w}{w} = \cfrac{1}{1 + \cfrac{w^2}{3 + \cfrac{w^2}{5 + \cfrac{w^2}{7 + \cfrac{w^2}{9 + \cdots}}}}}$$

expressed as a continued fraction and

$$= \frac{945 + 105w + w^2}{945 + 420w + 15w^2}$$

expressed as a ratio of polynomials reduced from the first five terms of the continued fraction. Using this and hyperbolic identies

$$e^{\frac{-qz}{2}} = \frac{1 - \tanh \frac{qz}{4}}{1 + \tanh \frac{qz}{4}} = \frac{1 - h\frac{qz}{4}}{1 + h\frac{qz}{4}}$$

and

$$\sinh \frac{qz}{2} = \frac{2 \tanh \frac{qz}{4}}{1 - \tanh^2 \frac{qz}{4}} = \frac{\frac{hqz}{2}}{1 - (\frac{hqz}{4})^2} .$$

Using all of these equations then

$$\Gamma(1-\alpha)\,[e^{-q(\alpha-1)} - 1] = \Gamma(2-\alpha)\, e^{\frac{-qz}{2}} \frac{\sinh \frac{qz}{2}}{\frac{qz}{2}} \cdot q$$

$$= \Gamma(2-\alpha) \frac{(1 - h\frac{qz}{4})}{(1 + h\frac{qz}{4})} \frac{h \frac{qz}{2}}{[1 - (\frac{hqz}{4})^2]} \cdot \frac{q}{\frac{qz}{2}} =$$

$$= \Gamma(2-\alpha) \frac{hq}{[1 + \frac{hqz}{4}]^2}$$

The properties of the model require that the parameter α be greater than or at least equal to one. Since the system is fairly unstable, it was found in actual calculation that the iterative Newton procedure described previously would sometimes give an estimate of α less than one. To avoid this difficulty a restriction was placed on the procedure as follows: Given that $\hat{\alpha}_i = 1 + \delta_i$ then

$$\hat{\alpha}_{i+1} = \begin{cases} \hat{\alpha}_{i+1} & \text{if } \hat{\alpha}_{i+1} \geq 1 + \frac{\delta_i}{2} \\ \frac{1+\delta_i}{2} & \text{if } \hat{\alpha}_{i+1} < 1 + \frac{\delta_i}{2} \end{cases}$$

If indeed $\alpha = 1$ it would be hoped that $\hat{\alpha}_i \to 1$ from above but this is not the case since the method blows up; that is for $\hat{\alpha}$ less than about 1.005 (depending on the data) the error in calculating $\Delta\alpha$ is larger than $\hat{\alpha}$ itself which reduces the iteration to nonsense. What results then is that the estimate is cut half way to one each time until which time the error in $\Delta\alpha$ causes a large positive jump. The estimate again approaches one and the process repeated until the computer is stopped by a programmed check which halts the Newton process after 50 iterations if no solution is reached. If this happens α is set equal to one in the original likelihood equation and another method used to estimate the other parameters.

Consider therefore the likelihood function

$$L(B,1,\eta) = \sum_{m=1}^{\infty} - B\eta^m \frac{\Gamma(m)}{m!} + \sum_{m=1}^{\infty} y_m \,[\log B + m \log \eta + \log\Gamma(m) - \log m!].$$

$$= -B\sum_{m=1}^{\infty} \frac{\eta^m}{m} + \sum_{m=1}^{\infty} y_m \,[\log B + m \log \eta + \log \Gamma(m) - \log m!]$$

Using the expansion

$$\log x = \frac{x-1}{x} + \frac{1}{2}\left(\frac{x-1}{x}\right)^2 + \frac{1}{3}\left(\frac{x-1}{x}\right)^3 + \ldots \text{ for } x > \tfrac{1}{2}$$

set $\eta = \frac{x-1}{x}$

$x\eta = x-1$

$x(1-\eta) = 1$

$x = \frac{1}{(1-\eta)}$ for $0 < \eta < 1$ then $1 < x < \infty$

thus

$$\sum_{m=1}^{\infty} \frac{\eta^m}{m} = \log x = \log\left(\frac{1}{1-\eta}\right) = -\log(1-\eta).$$ Using this and

again making the substitution $(1-\eta) = e^{-q}$

$$L(B,1,q) = -Bq + \sum_{m=1}^{\infty} y_m \left[\log B + m \log(1-e^{-q}) + \log \Gamma(m) - \log m!\right].$$

Taking the derivatives with respect to these parameters

$$L_B = -q + \sum_{m=1}^{\infty} y_m/B = -q + \frac{s}{B}$$

$$L_q = -B + \sum_{m=1}^{\infty} y_m \frac{m\, e^{-q}}{1-e^{-q}} = -B + I\, \frac{1}{e^q - 1}.$$

In finding a solution $L_B = L_q = 0$ make the substitution $B = \frac{s}{q}$ into the second equation to get

$$-\frac{s}{q} + I\, \frac{1}{e^q - 1} = 0$$

which reduces to $e^q - 1 - \frac{I}{s} q = 0$.

To find a solution to this equation consider the following iterative procedure. Given an initial estimate q_0 and where q_r

is the root of the equation

$$q_r = q_0 + \frac{x}{1+ax+bx^2} = q_0 + \frac{x}{B(x)}$$ where x, a, b are to be determined as follows:

Let $\lambda = \frac{I}{s}$ and $A_0^* = e^{q_0} - 1 - \lambda q_0$

Set $e^{q_r} - 1 - \lambda q_r = e^{q_0 + \frac{x}{B(x)}} - 1 - \lambda(q_0 + \frac{x}{B(x)}) = 0.$

Then

$$B(x)\,[e^{q_0} e^{\frac{x}{B(x)}} - 1 - \lambda q_0 - \lambda \frac{x}{B(x)}] = 0.$$

Expanding this equation and calling it Q(x) then

$$Q(x) = B(x)e^{q_0}[1 + \frac{x}{B(x)} + \frac{x^2}{2B^2(x)} + \frac{x^3}{6B^3(x)} + \frac{x^4}{24B^4(x)} + \ldots.]$$

$$-B(x) - \lambda B(x)q_0 - \lambda x$$

$$= B(x)\,A_0^* + (e^{q_0} - \lambda)x + e^{q_0}[\frac{x^2}{2B(x)} + \frac{x^3}{6B^2(x)} + \frac{x^4}{24B^3(x)} + \ldots.].$$

Now expand the expressions $\frac{1}{B^k(x)} = \left(\frac{1}{1+ax+bx^2}\right)^k$ into polynomials

$$\left(\frac{1}{1+ax+bx^2}\right)^k = a_{k0} + a_{k1}x + a_{k2}x^2 + a_{k3}x^3 + \ldots.$$

After finding these polynomials for k = 1,2,3 the expansion then becomes Q(x) =

$$= B(x)A^*_0 + (e^{q_0} - \lambda)x + e^{q_0}\,[\tfrac{1}{2}x^2(1 - ax + (a^2 - b)x^2 + \ldots)$$

$$+ \frac{1}{6}x^3 (1-2ax + \ldots) + \frac{1}{24}x^4 (1 + \ldots.)] + \ldots.$$

$$+ B(x) A_0^* + (e^{q_0} - \lambda)x + \frac{e^{q_0}}{2} x^2 + e^{q_0} x^3 [\tfrac{1}{6} - \tfrac{1}{2}a]$$

$$+ e^{q_0} x^4[\tfrac{1}{24} - \tfrac{a}{3} + \tfrac{1}{2} (a^2-b)] + \ldots.$$

Choose a and b such that the coefficient of x^3 and x^4 are zero. Thus

$$\frac{1}{6} - \frac{1}{2}a = 0 \Rightarrow a = \frac{1}{3}$$

and

$$\frac{1}{24} - \frac{a}{3} + \frac{1}{2} (a^2 - b) = 0$$

$$\frac{1}{24} - \frac{1}{9} + \frac{1}{2} (\frac{1}{9} - b) = 0 \Rightarrow b = - \frac{1}{36}$$

therefore

$$Q(x) = A_0^* (1 + \tfrac{1}{3}x - \tfrac{1}{36}x^2) + (e^{q_0} - \lambda)x + \frac{e^{q_0}}{2} x^2 + a_5x^5 + \ldots.$$

$$= A_0^* + (e^{q_0} - \lambda + \frac{A_0^*}{3})x + (\frac{e^{q_0}}{2} - \frac{A_0^*}{36})x^2 + a_5x^5 + \ldots.$$

As an approximation to $Q(x) = 0$ set the equation

$$A_0^* + (e^{q_0} - \lambda + \frac{A_0^*}{3})x + (\frac{e^{q_0}}{2} - \frac{A_0^*}{36})x^2 = 0$$

and solve for x as follows: For the general quadratic
$\alpha x^2 + \beta x + \delta = 0$

$$x = \frac{-\beta + \sqrt{\beta^2 - 4\alpha\delta}}{2\alpha} \frac{(-\beta - \sqrt{\beta^2 - 4\alpha\delta})}{(-\beta - \sqrt{\beta^2 - 4\alpha\delta})} = \frac{2\delta}{-\beta - \sqrt{\beta^2 - 4\alpha\delta}}$$

Since β is positive in the neighborhood of q_r the positive root is taken to obtain the root of the quadratic nearer zero and the last form is used in calculating x to avoid round off. The procedure for finding the root of the equation $e^q - 1 - \lambda q = 0$ is as follows:

1. Make initial estimate $q_0 = \log(1+\lambda \log \lambda)$
2. Evaluate $A_i^* = e^{q_i} - 1 - \lambda q_i$
3. Solve the equation
$$A_i^* + (e^{q_i} - \lambda + \frac{A_i^*}{3})x + (\frac{e^{q_i}}{2} - \frac{A_i^*}{36})x^2 = 0 \text{ for } x.$$
4. $q_{i+1} = q_i + \dfrac{x}{1+\frac{1}{3}x-\frac{1}{36}x^2}$
5. Return to step 2 until desired accuracy is reached.

The method was designed for rapid convergence and in fact it was found in actual compitation that five digit accuracy was obtained in only two iterations.

Using $\hat{q}$ as found from the above procedure and from the original equations remembering that A = B for the case in question where $\alpha = 1$ the estimates obtained are

$$\hat{A} = \frac{s}{\hat{q}}$$

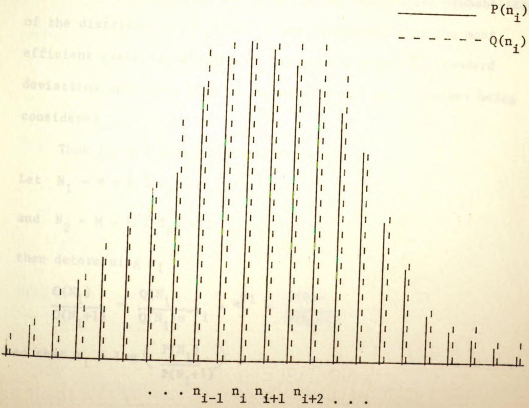
and

$$\hat{k}_s = \frac{\eta}{1-\eta} = \frac{1-e^{-\hat{q}}}{e^{-\hat{q}}} = e^{\hat{q}}-1.$$

Chapter 3 Generating Random Variables for the Simulation

Section 1: Acceptance Rejection Procedures

In the process of simulating the established model on the high speed computer it is necessary to generate random variables with certain specified distributions. Since the actual computation is to be done on the computer and the procedures used many thousands of times it is necessary they be effecient and use the minimum of input random variables. With these goals in mind it was decided that for discrete random variables an acceptance rejection procedure would be used. This method of generating random variables with specified distributions is discussed by Rubin [5] and will be used in this problem in the following way:



Suppose a random variable with the distribution $P(n_i)$ is desired. Construct a frequency distribution $Q(n_i)$ which dominates $P(n_i)$. Obtain an observation from the distribution $Q(n_i)$ and accept this observation x_1 with probability $\frac{P(x_1)}{Q(x_1)}$. If x_1 is rejected obtain a second observation from $Q(n_i)$ and repeat the process until an observation is accepted. If the first accepted observation is designated as x then it has distribution $P(n_i)$.

This procedure is to be used for Binomial, Poisson and Hypergeometric distributions and in these cases the distribution $Q(n_i)$ will take the form of a uniform over the mode and discrete exponential over each tail with parameter α_1 over the right tail and parameter α_2 over the left tail. The parameters α_1 and α_2 are determined by taking the ratio of two consecutive probabilities of the distribution $P(n_i)$ and it was determined that the most efficient place to calculate α_1 and α_2 was about $\sqrt{2}$ standard deviations on either side of the mode of the distributions being considered.

Thus let M = mode of distribution $P(n_i)$

Let $N_1 = M + [\sqrt{2}\,\sigma_p]$

and $N_2 = M - [\sqrt{2}\,\sigma_p] - 1$

then determining α_1

$$\frac{Q(N_1)}{Q(N_1+1)} = \frac{Q(N_1)}{Q(N_1)e^{-\alpha_1}} = e^{\alpha_1} = \frac{P(N_1)}{P(N_1+1)}$$

so that $\alpha_1 = \log\left(\frac{P(N_1)}{P(N_1+1)}\right)$

and similarly for α_2

$$\frac{Q(N_2+1)}{Q(N_2)} = \frac{Q(N_2+1)}{Q(N_2+1)e^{-\alpha_2}} = e^{\alpha_2} = \frac{P(N_2+1)}{P(N_2)}$$

so that $\alpha_2 = \log\left(\frac{P(N_2+1)}{P(N_2)}\right)$.

The first term of the right exponential is N_1-k where

$$k = \left[\frac{\log P(M) - \log P(N_1)}{\alpha_1}\right]$$

and

$$\sum_{i=0}^{\infty} Q(N_1-k+i) = \sum_{i=0}^{\infty} P(N_1)\, e^{\alpha_1(k-i)}$$

$$= P(N_1)\, e^{k\alpha_1} \sum_{i=0}^{\infty} e^{-i\alpha_1} = \frac{P(N_1)e^{k\alpha_1}}{1-e^{-\alpha_1}}$$

also the last term of the left exponential is

$N_2 + 1 + j$ where

$$j = \left[\frac{\log P(M) - \log P(N_2+1)}{\alpha_2}\right]$$

and

$$\sum_{i=0}^{\infty} Q(N_2 + 1 + j - i) = \sum_{i=0}^{\infty} P(N_2+1)\, e^{\alpha_2(j-i)}$$

$$= P(N_2+1)\, e^{j\alpha_2} \sum_{i=0}^{\infty} e^{-i\alpha_2} = \frac{P(N_2+1)e^{j\alpha_2}}{1-e^{-\alpha_2}}$$

$Q(i)$ being thus defined in the tails let

$$Q(i) = P(M) \quad \text{for } N_2 + 1 + j < i < N_1 - k$$

so that

$$\sum_{i=-\infty}^{\infty} Q(i) = \frac{P(N_2+1)e^{j\alpha_2}}{1 - e^{-\alpha_2}} + P(M)\,(N_1-N_2-k-j-2) + \frac{P(N_1)e^{k\alpha_1}}{1-e^{-\alpha_1}}$$

For ease of computation make the substitutions

$$u = \log P(M) - \log P(N_1) - k\alpha_1$$

$$v = \log P(M) - \log P(N_2+1) - j\alpha_2$$

which reduces the sum to

$$\sum_{i=-\infty}^{\infty} Q(i) = P(M)\left[\frac{e^{-v}}{1-e^{-\alpha_2}} + (N_1-N_2-k-j-2) + \frac{e^{-u}}{1-e^{-\alpha_1}}\right].$$

By letting $T = \sum_{i=-\infty}^{\infty} Q(i) \,/\, P(M) =$

$$= \left[\frac{e^{-v}}{1-e^{-\alpha_2}} + (N_1 - N_2 - k - j - 2) + \frac{e^{-u}}{1-e^{-\alpha_1}}\right]$$

and normalizing this quantity a random variable with the distribution P(i) can be found as follows:

1) Let U_1 be a uniform random variable

2) If $U_1 < \dfrac{e^{-v}}{T(1-e^{-\alpha_2})}$ the observation is to be taken from the left tail. Thus choose $N_0 = \left[-\frac{1}{\alpha_2}\log U_{11}\right]$ where U_{11} is a uniform random variable and the brackets indicates the greatest integer contained in the bracketed quantity. The observation thus becomes $N = N_2 + 1 + j - N_0$. Then accept N with probability $\dfrac{P(N)}{Q(N)}$.

3) If $\dfrac{e^{-v}}{T(1-e^{-\alpha_2})} \leq U_1 \leq 1 - \dfrac{e^{-u}}{T(1-e^{-\alpha_1})}$ the observation is to be taken from the uniform range as follows

$$R = \frac{U_1 - \dfrac{e^{-v}}{T(1-e^{-\alpha_2})}}{1 - \dfrac{e^{-u}}{T(1-e^{-\alpha_1})} - \dfrac{e^{-v}}{T(1-e^{-\alpha_2})}} \quad (N_1 - N_2 - k - j - 2)$$

and let $N_0 = [R]$

so that the observation N is

$$N = N_2 + j + 2 + N_0$$

and N is accepted with probability $\frac{P(N)}{Q(N)}$.

4) If $U_1 > 1 - \frac{e^{-u}}{T(1-e^{-\alpha_1})}$ the observation is taken from the right tail using the same procedure as in step 2. That is choose $N_0 = \left[-\frac{1}{\alpha_1} \log U_{12} \right]$ where U_{12} is again uniform only this time let the observation be

$N = N_1 - k + N_0$ and accept N with probability $\frac{P(N)}{Q(N)}$.

5) If at step 2,3 or 4,N is rejected obtain a new uniform U_2 and repeat the process until an observation N is accepted. N will then be distributed according to the distribution $P(n_i)$.

In steps 2,3,4 the acceptance rejection part of the procedure is handled in comparison with an exponential random variable E_0 in the following way: Accept N if

$$E_0 \geq -\log \frac{P(N)}{Q(N)} = \log Q(N) - \log P(N).$$

Where in the left tail

$$\log Q(N) = \log P(N_2 + 1) + (j - N_0)\,\alpha_2$$

in the right tail

$$\log Q(N) = \log P(N_1) + (k - N_0)\,\alpha_1$$

and in the uniform range

$$\log Q(N) = \log P(M).$$

This method of comparison is used so it is not necessary to calculate $Q(M)$, $Q(N_2 + 1)$ and $Q(N_1)$ using instead already calculated quantities.

Section 2: Fitting Discrete Distributions with Large Means

In the problem at hand the Poisson, Binomial and Hypergeometric distributions are used. It is therefore necessary to determine the distribution $Q(n_i)$ as developed in the previous section for these cases but first it will be necessary to develop some machinery for the calculation of $\log x!$ which is necessary to evaluate in all three of the above mentioned distributions when calculating $\log P(n_i)$.

The first equation used is Stirling's asymptotic approximation to n!. From this

$$\log n! = (n + \tfrac{1}{2}) \log n - n + \tfrac{1}{2} \log 2\pi + \varphi(n)$$

where $\varphi(n) = \frac{1}{12n} - \frac{1}{360n^3} + \frac{1}{1260n^5} - \frac{1}{1680n^7} \pm \cdots$.

Consider now the product

$$2^{2n}\,\Gamma(n+1)\,\Gamma(n+\tfrac{1}{2}) = n!\ 1\cdot3\cdot5\cdots(2n-3)(2n-1)\,\frac{\sqrt{\pi}}{2^n}\,2^{2n}$$

$$= \sqrt{\pi}\,\Gamma(2n+1).$$

Taking the logarithm of both sides

$$\tfrac{1}{2}\log\pi + \log\Gamma(2n+1) = 2n\log 2 + \log\Gamma(n+\tfrac{1}{2}) + \log n!$$

and using the more general form of Stirling's equation

$$\log\Gamma(x+1) = (x+\tfrac{1}{2})\log x - x + \tfrac{1}{2}\log 2\pi + \varphi(x)$$

where $\varphi(x) = \sum_{m=0}^{\infty} C_m x^{-m}$.

Thus

$$\tfrac{1}{2}\log\pi + \log\Gamma(2n+1) = \tfrac{1}{2}\log\pi + \log\Gamma(2n+2) - \log(2n+1)$$

$$= \tfrac{1}{2}\log\pi + (2n+\tfrac{1}{2})\log(2n+1) - (2n+1) + \tfrac{1}{2}\log 2\pi + \sum_{m=0}^{\infty} C_m (2n+1)^{-m}.$$

Also $\tfrac{1}{2}\log\pi + \log\Gamma(2n+1) =$

$$= 2n\log 2 + \log\Gamma(n+3/2) - \log(n+\tfrac{1}{2}) + \log n!$$

$$= 2n\log 2 + n\log(n+\tfrac{1}{2}) - (n+\tfrac{1}{2}) + \tfrac{1}{2}\log 2\pi + \sum_{m=0}^{\infty} C_m (n+\tfrac{1}{2})^{-m} + \log n!.$$

Collecting terms then and combining these two equations

$$\log n! = \tfrac{1}{2}\log\pi + (2n+\tfrac{1}{2})\log(2n+1) - (2n+\tfrac{1}{2})\log 2 + \tfrac{1}{2}\log 2$$
$$- n\log(n+\tfrac{1}{2}) - (n+\tfrac{1}{2}) - \sum_{m=0}^{\infty} C_m (n+\tfrac{1}{2})^{-m} + \sum_{m=0}^{\infty} \frac{C_m}{2^m}(n+\tfrac{1}{2})^{-m}$$
$$= \tfrac{1}{2}\log 2\pi + (n+\tfrac{1}{2})\log(n+\tfrac{1}{2}) - (n+\tfrac{1}{2}) - \Psi(n+\tfrac{1}{2})$$

where

$$\Psi(n+\tfrac{1}{2}) = \sum_{m=0}^{\infty} C_m (n+\tfrac{1}{2})^{-m}(1-2^{-m}).$$

Note that this function Ψ is not the logarithmic derivative of the gamma function used in Chapter 2.

Since the original equation for log n! was an asymptotic approximation, log n! and therefore $\Psi(n+\frac{1}{2})$ cannot be calculated in this way for small n. To find $\Psi(n+\frac{1}{2})$ for $n = 0,1,\ldots 10$ calculate $\Psi(11+\frac{1}{2}) = \Psi(11.5)$ from the already developed formula and use a backwards recursion formula which is now to be derived.

$$\log n! = \tfrac{1}{2}\log 2\pi + (n+\tfrac{1}{2})\log(n+\tfrac{1}{2}) - (n+\tfrac{1}{2}) - \Psi(n+\tfrac{1}{2})$$
$$= \tfrac{1}{2}\log 2\pi + (n+\tfrac{1}{2})\log(1+\tfrac{1}{2n}) + (n+\tfrac{1}{2})\log n - (n+\tfrac{1}{2}) - \Psi(n+\tfrac{1}{2})$$

also

$$\log n! = \log n + \log(n-1)!$$
$$= \log n + \tfrac{1}{2}\log 2\pi + (n-\tfrac{1}{2})\log(1-\tfrac{1}{2n}) + (n-\tfrac{1}{2})\log n - (n-\tfrac{1}{2}) - \Psi(n-\tfrac{1}{2}).$$

Combining these equations

$$\Psi(n-\tfrac{1}{2}) = \Psi(n+\tfrac{1}{2}) + 1 + (n-\tfrac{1}{2})\log(1-\tfrac{1}{2n}) - (n+\tfrac{1}{2})\log(1+\tfrac{1}{2n})$$
$$= \Psi(n+\tfrac{1}{2}) + 1 + (n-\tfrac{1}{2})\left[-\frac{1}{2n} - \frac{1}{2}\frac{1}{4n^2} - \frac{1}{3}\frac{1}{8n^3} - \frac{1}{4}\frac{1}{16n^4} - \frac{1}{5}\frac{1}{2^5n^5} - \cdots\right]$$
$$- (n+\tfrac{1}{2})\left[\frac{1}{2n} - \frac{1}{2}\frac{1}{4n^2} + \frac{1}{3}\frac{1}{2^3n^3} - \frac{1}{4}\frac{1}{2^4n^4} + \frac{1}{5}\frac{1}{2^5n^5} \mp \cdots\right]$$
$$= \Psi(n+\tfrac{1}{2}) + \frac{1}{2^2\cdot 2\cdot 3n^2} + \frac{1}{2^4\cdot 4\cdot 5n^4} + \frac{1}{2^6\cdot 6\cdot 7\cdot n^6} + \cdots \frac{1}{2^{2k}\cdot 2k(2k+1)n^{2k}} + \cdots$$

The first seven terms of this expansion are used for n = 1, 2,...,10 but in calculating $\Psi(\frac{1}{2})$ an additional four terms are used.

Consider now a careful calculation of the expression $(1+x)\log(1+x)-x$ which will be useful in calculating $\log P(n_i)$. Make the substitution $1+x = \frac{1+y}{1-y}$ so that

$$x = \frac{2y}{1-y} \qquad \text{and} \qquad y = \frac{x}{2+x}$$

and under the assumption that $x > -1$ it follows that $|y| < 1$.

Thus $(1+x)\log(1+x) - x = \frac{1+y}{1-y}\log\left(\frac{1+y}{1-y}\right) - \frac{2y}{1-y}$.

The evaluation of this expression will be broken into two cases

First if $\frac{1}{2} < |y| < 1$ then

$$(1+x)\log(1+x) - x = (1+x)\log(1+x) - \frac{x}{(2+x)}(2+2x-x)$$

$$= xy + (1+x)[\log(1+x) - 2y].$$

Secondly if $|y| \leq \frac{1}{2}$ use the expansion

$$\log\left(\frac{1+y}{1-y}\right) = 2y + \frac{2}{3}y^3 + \frac{2}{5}y^5 + \ldots.$$

Thus

$$(1+x)\log(1+x) - x$$

$$= \left(\frac{1+y}{1-y}\right)[2y + \frac{2}{3}y^3 + \frac{2}{5}y^5 + \ldots] - \frac{2y}{1-y}$$

$$= \frac{2y^2}{1-y} + \frac{1+y}{1-y}[\frac{2}{3}y^3 + \frac{2}{5}y^5 + \frac{2}{7}y^7 + \frac{2}{9}y^9 + \ldots]$$

$$= xy + (1+x)[\frac{2}{3}y^3 + \frac{2}{5}y^5 + \frac{2}{7}y^7 + \frac{2}{9}y^9 + \ldots].$$

With these equations consider now the calculation of $Q(n_i)$ for desired distributions.

1) Poisson Distribution: Let λ be the parameter of the Poisson distribution.

Then obviously $M = [\lambda]$

$$N_1 = M + [\sqrt{2\lambda}]$$

$$N_2 = M - [\sqrt{2\lambda}] - 1$$

where in each case the bracket indicates the greatest integer contained in the bracket.

Also

$$\frac{Q(N_1)}{Q(N_1+1)} = e^{\alpha_1} = \frac{P(N_1)}{P(N_1+1)} = \frac{N_1+1}{\lambda}$$

so that $\alpha_1 = \log(N_1+1) - \log\lambda$.

Similarly

$$\frac{Q(N_2+1)}{Q(N_2)} = e^{\alpha_2} = \frac{P(N_2+1)}{P(N_2)} = \frac{\lambda}{N_2+1}$$

so that $\alpha_2 = \log\lambda - \log(N_2+1)$.

and finally $P(n) = \frac{\lambda^n}{n!} e^{-\lambda}$ for $n = 0,1,2,3,4,\ldots$

so $\log P(n) = -\lambda + n\log\lambda - \log n!$

$= -\lambda + n\log\lambda - (n+\frac{1}{2})\log(n+\frac{1}{2}) + (n+\frac{1}{2}) - \frac{1}{2}\log 2\pi + \Psi(n+\frac{1}{2})$.

Make the substitution $n = \lambda + \mu$. Then

$$\log P(n) = -\lambda + (\lambda+\mu+\tfrac{1}{2})\log\lambda - (\lambda+\mu+\tfrac{1}{2})\log(\lambda+\mu+\tfrac{1}{2}) - (\lambda+\mu+\tfrac{1}{2}) + \tfrac{1}{2}\log 2\pi - \Psi(\lambda+\mu+\tfrac{1}{2}) - \tfrac{1}{2}\log\lambda$$

$$= -\lambda[1 + \frac{\mu+\frac{1}{2}}{\lambda}]\log[1 + \frac{\mu+\frac{1}{2}}{\lambda}] + \lambda[\frac{\mu+\frac{1}{2}}{\lambda}] - \tfrac{1}{2}\log 2\pi\lambda + \Psi(\lambda+\mu+\tfrac{1}{2})$$

$$= -\lambda\left\{[1 + \frac{\mu+\frac{1}{2}}{\lambda}]\log[1 + \frac{\mu+\frac{1}{2}}{\lambda}] - \frac{\mu+\frac{1}{2}}{\lambda}\right\} - \tfrac{1}{2}\log 2\pi\lambda + \Psi(\lambda+\mu+\tfrac{1}{2}).$$

2) Binomial Distribution: Let p,n be the parameters of the Binomial distribution

then $M = [(n+1)p]$

$$N_1 = M + [\sqrt{2npq}]$$

$$N_2 = M - [\sqrt{2npq}] - 1$$

and as before $e^{\alpha_1} = \frac{N_1+1}{n-N_1} \frac{1-p}{p}$ and $e^{\alpha_2} = \frac{n-N_2}{N_2+1} \frac{p}{1-p}$

so $\alpha_1 = \log \frac{N_1+1}{n-N_1} + \log \frac{1-p}{p}$ and $\alpha_2 = \log \frac{n-N_2}{N_2+1} - \log \frac{1-p}{p}$

and finally $P(k) = \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k}$ for $k = o,1,\ldots,n$.

$$\log P(k) = \log n! + k \log p + (n-k) \log(1-p) - \log k! - \log(n-k)! .$$

Using the derived formula for log x! then and the identity $\log n! = \log(n+1)! - \log(n+1)$

$$\begin{aligned}\log P(k) = {} & (n+\tfrac{1}{2}) \log(n+1) - (n+1) + \tfrac{1}{2} \log 2\pi + \varphi(n+1) \\ & - [(k+\tfrac{1}{2}) \log(k+\tfrac{1}{2}) - (k+\tfrac{1}{2}) + \tfrac{1}{2} \log 2\pi - \Psi(k+\tfrac{1}{2})] \\ & - [(n-k+\tfrac{1}{2}) \log(n-k+\tfrac{1}{2}) - (n-k+\tfrac{1}{2}) + \tfrac{1}{2}\log 2\pi - \Psi(n-k+\tfrac{1}{2})] \\ & + k \log p + (n-k) \log(1-p).\end{aligned}$$

Make the substitution $k + \frac{1}{2} = (n+1)p+\mu$ into the above equation.

$$\begin{aligned}\log P(k) = {} & (n+\tfrac{1}{2}) \log(n+1) - [(n+1)p+\mu] \log((n+1)p+\mu) \\ & - [(n+1)q-\mu] \log((n+1)q-\mu) + [(n+1)p+\mu] \log p \\ & + [(n+1)q-\mu] \log q - \tfrac{1}{2} \log 2\pi + \varphi(n+1) + \Psi(k+\tfrac{1}{2}) + \Psi(n-k+\tfrac{1}{2}) \\ & - \tfrac{1}{2} \log p - \tfrac{1}{2} \log q \\ = {} & - [(n+1)p+\mu] \log[1 + \frac{\mu}{(n+1)p}] - [(n+1)p+\mu] \log(n+1)p \\ & - [(n+1)q-\mu] \log[1 - \frac{\mu}{(n+1)q}] - [(n+1)q-\mu] \log(n+1)q \\ & + [(n+1)q-\mu] \log q + [(n+1)p+\mu] \log p + (n+1) \log(n+1) \\ & - \tfrac{1}{2} \log (n+1) 2\pi pq + \varphi(n+1) + \Psi(k+\tfrac{1}{2}) + \Psi(n-k+\tfrac{1}{2}) \\ = {} & - (n+1) p\left\{[1 + \frac{\mu}{(n+1)p}] \log(1 + \frac{\mu}{(n+1)p}) - \frac{\mu}{(n+1)p}\right\} \\ & - (n+1)q \left\{[1 - \frac{\mu}{(n+1)q}] \log(1 - \frac{\mu}{(n+1)q}) - \left(\frac{-\mu}{(n+1)q} \right)\right\} \\ & - \tfrac{1}{2} \log(n+1)2\pi pq + \varphi(n+1) + \Psi(k+\tfrac{1}{2}) + \Psi(n-k+\tfrac{1}{2}) .\end{aligned}$$

3) Hypergeometric Distribution: Let D,N,n be the parameters of the distribution

then

$$M = \left[\frac{(n+1)(D+1)}{N+2}\right]$$

$$N_1 = M + \left[\frac{nD(N-D)(N-n)}{N^2(N-1)}\right]$$

$$N_2 = M - \left[\frac{nD(N-D)(N-n)}{N^2(N-1)}\right] - 1$$

also

$$e^{\alpha_1} = \frac{(N_1+1)(N-D-n+N_1+1)}{(D-N_1)\ (n-N_1)} \quad \text{and} \quad e^{\alpha_2} = \frac{(D-N_2)(n-N_2)}{(N_2+1)(N-D-n+N_2+1)}.$$

Consider now the probability

$$P(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} = \frac{D!}{(D-x)!x!}\ \frac{(N-D)!}{(n-x)!(n-D-n+x)!}\ \frac{n!(N-n)!}{N!}$$

$$= \frac{C(N,n,D)}{(D-x)!\ x!\ (n-x)!\ (N-D-n+x)!} \quad \text{for } x = 0,1,\ldots,D.$$

Expand the factorials using the established formula and make the substitution

$$y = x + \tfrac{1}{2} - M_0 \qquad \text{where } M_0 = \frac{(n+1)(D+1)}{N+2}.$$

Thus

$$\log x! = (x+\tfrac{1}{2})\log(x+\tfrac{1}{2}) - (x+\tfrac{1}{2}) + \tfrac{1}{2}\log 2\pi - \Psi(x+\tfrac{1}{2})$$

$$= (M_0+y)\log(M_0+y) - (M_0+y) + \tfrac{1}{2}\log 2\pi - \Psi(M_0+y)$$

$$= M_0\left[(1+\frac{y}{M_0})\log(1+\frac{y}{M_0}) - \frac{y}{M_0}\right] + y\log M_0 - M_0 + M_0\log M_0$$

$$+ \tfrac{1}{2}\log 2\pi - \Psi(M_0+y)$$

$$\log(n-x)! = (n-x+\tfrac{1}{2})\log(n-x+\tfrac{1}{2}) + \tfrac{1}{2}\log 2\pi - \Psi(n-x+\tfrac{1}{2})$$

$$= (n+1-M_0-y)\log(n+1-M_0-y) - (n+1-M_0-y) + \tfrac{1}{2}\log 2\pi - \Psi(n+1-M_0-y)$$

$$= (n+1-M_0)\left[(1-\frac{y}{n+1-M_0})\log(1-\frac{y}{n+1-M_0}) - \frac{-y}{n+1-M_0}\right] - y\log(n+1-M_0)$$

$$+ (n+1-M_0)\log(n+1-M_0) - (n+1-M_0) + \tfrac{1}{2}\log 2\pi - \Psi(n+1-M_0-y).$$

Similarly as above

$$\log(D-x)! = (D+1-M_0)\left[(1-\frac{y}{D+1-M_0})\log(1-\frac{y}{D+1-M_0}) - \frac{-y}{D+1-M_0}\right]$$

$$- y\log(D+1-M_0) + (D+1-M_0)\log(D+1-M_0) - (D+1-M_0)$$

$$+ \tfrac{1}{2}\log 2\pi - \Psi(D+1-M_0-y)$$

and finally

$$\log(N-D-n+x)! = (N-D-n+M_0)\left[(1+\frac{y}{N-D-n+M_0})\log(1+\frac{y}{N-D-n+M_0})\right.$$

$$\left.-\frac{y}{N-D-n+M_0}\right] + y\log(N-D-n+M_0)$$

$$+ (N-D-n+M_0)\log(N-D-n+M_0) - (N-D-n+M_0)$$

$$+ \tfrac{1}{2}\log 2\pi - \Psi(N-D-n+M_0+y).$$

Combining these equations

$$\log P(x) = C^*(N,n,D) - M_0\left[(1+\frac{y}{M_0})\log(1+\frac{y}{M_0}) - \frac{y}{M_0}\right]$$

$$- (n+1-M_0)\left[(1-\frac{y}{n+1-M_0})\log(1-\frac{y}{n+1-M_0}) - \frac{-y}{n+1-M_0}\right]$$

$$- (D+1-M_0)\left[(1-\frac{y}{D+1-M_0})\log(1-\frac{y}{D+1-M_0}) - \frac{-y}{D+1-M_0}\right]$$

$$- (N-D-n+M_0)\left[(1+\frac{y}{N-D-n+M_0})\log(1+\frac{y}{N-D-n+M_0}) - \frac{y}{N-D-n+M_0}\right]$$

$$- y\log\left[\frac{M_0(N-D-n+M_0)}{(n+1-M_0)(D+1-M_0)}\right]$$

$$+ \Psi(M_0+y) + \Psi(n+1-M_0-y) + \Psi(D+1-M_0-y) + \Psi(N-D-n+M_0+y).$$

A check will show that $\frac{M_0(N-D-n+M_0)}{(n+1-M_0)(D+1-M_0)} = 1$ which eliminates this term from consideration.

Also for use in these acceptance rejection procedures the constant term $C^*(N,n,D)$ may be neglected since the procedure uses only the ratios of the probabilities of the respective points being considered.

Section 3: Procedures for Discrete Distributions with Small Means.

The procedures in the previous section generate the desired random variable using a small number of uniform random variables but at the expense of considerable numerical calculations. When the mean of the distributions under consideration is small, procedures exist which use about the same number of uniform random variables but which are much less involved. Such procedures used in the simulation will now be considered.

1) Poisson: Let λ be the mean of the Poisson distribution. Let $E_1, E_2, E_3, \ldots$ be independent exponential random variables which are obtained by the equation $E_i = -\log U_i$ where U_i are independent uniform random variables. Let J be the integer such that

$$\sum_{i=1}^{J-1} E_i < \lambda \le \sum_{i=1}^{J} E_i$$

Then J-1 has a Poisson distribution with mean λ and $\sum_{i=1}^{J} E_i - \lambda$ is independent exponential. This result can be shown by directly integrating the joint density of the E_i.

2) Truncated Poisson: This distribution is needed only in the small mean case and its use will be shown later. Let λ be the mean of the distribution and as before let $E_1, E_2, E_3, \ldots$ be exponential random variables.

Let q be defined as the integer such that

$$q\lambda < E_1 \le (q+1)\lambda .$$

Let J be the integer such that

$$\sum_{i=1}^{J-1} E_i < (q+1)\lambda \le \sum_{i=1}^{J} E_i .$$

Then J-1 has a truncated Poisson distribution with mean λ and $\sum_{i=1}^{J} E_i - (q+1)\lambda$ is independent exponential. This result can also be shown by directly integrating the joint density of the E_i.

3) Binomial: Let N,p be the parameters of the Binomial distribution.

Define $\alpha = -\log(1-p)$ and let $g = N\alpha$. Divide the interval $(0,N\alpha]$ into the N intervals $I_i = ((i-1)\alpha, i\alpha]$.

Let $E_1, E_2, E_3, \ldots$ be independent exponential random variables. Consider the points

$$x_i = \sum_{j=1}^{i} E_j \quad \text{for } i = 1,2,3,\ldots,k-1$$

where k is defined to be the first integer such that

$$x_k = \sum_{j=1}^{k} E_j > N\alpha.$$

Let N_B = Number of intervals I_i which contain a point x_i.

Then N_B has a Binomial distribution with parameters N and p. This can be shown directly by integrating the joint density of the E_j.

4) Hypergeometric: Let N,D and n be the parameters of the distribution. Then

$$P(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} = \binom{D}{x}\frac{n!}{(N-n)!N!} \quad \frac{(N-D)!}{(N-D-n+x)!(n-x)!}$$

$$= \frac{(N-D)!}{(N-D-n)!(N-n)!N!} \binom{D}{x}\left(\frac{p^*}{1-p^*}\right)^x \left[\frac{n!(N-D-n)!}{(N-D-n+x)!(n-x)!} \left(\frac{1-p^*}{p^*}\right)^x\right]$$

where $\frac{p^*}{1-p^*} = \frac{n}{N-D-n+1}$ and consequently $p^* = \dfrac{\frac{n}{N-D-n+1}}{1+\frac{n}{N-D-n+1}}$.

Let $N_1 \sim B(D;p^*)$ and accept N_1 with probability

$$\frac{n!(N-D-n)!}{(N-D-n+N_1)!\ (n-N_1)!} \left(\frac{N-D-n+1}{n}\right)^{N_1}.$$

If N_1 is rejected let $N_2 \sim B(D,p^*)$ and repeat the process. Let N_H be the first accepted N_i. Then N_H has a Hypergeometric distribution with parameters N,D,n. This procedure is used for small mean Hypergeometric and it is to be noted that for the case where x = 0,1 the acceptance probability is one so that the acceptance rejection part of the procedure is ignored when the Binomial random variable is zero or one.

Consider now a simplification of the factor,

$$R(x) = \frac{n!(N-D-n)!}{(N-D-n+x)!(n-x)!} \left(\frac{N-D-n+1}{n}\right)^{x}.$$

Using the established formula for log x!

$$\log n! = \log n + \log(n-1)! = (n+\tfrac{1}{2})\log n - n + \tfrac{1}{2}\log 2\pi + \varphi(n)$$

$$\log(n-x)! = (n-x+\tfrac{1}{2})\log(n-x+\tfrac{1}{2}) - (n-x+\tfrac{1}{2}) + \tfrac{1}{2}\log 2\pi - \Psi(n-x+\tfrac{1}{2})$$

Combining these with x log n

$$\log n! - \log(n-x) - x\log n = -(n-x+\tfrac{1}{2})\log(1-\frac{x-\frac{1}{2}}{n})$$

$$-(x-\tfrac{1}{2}) + \varphi(n) + \Psi(n-x+\tfrac{1}{2}).$$

Make the substitution $\mu = x-\frac{1}{2}$

$$= -(n-\mu)\log(1-\frac{\mu}{n}) - \mu + \varphi(n) + \Psi(n-\mu).$$

Similarly as above

$$\log(N-D-n)! - \log(N-D-n+x)! + x\log(N-D-n+1)$$

$$= -(N-D-n+1+\mu)\log(1+\frac{\mu}{N-D-n+1}) + \mu + \varphi(N-D-n+1) + \Psi(N-D-n+1+\mu).$$

From this then

$$\log R(\mu) = -n\left[(1-\frac{\mu}{n})\log(1-\frac{\mu}{n}) - \frac{-\mu}{n}\right]$$

$$-(N-D-n+1)\left[(1+\frac{\mu}{N-D-n+1})\log(1+\frac{\mu}{N-D-n+1}) - \frac{\mu}{N-D-n+1}\right]$$

$$+\varphi(n) + \Psi(n-\mu) + \varphi(N-D-n+1) + \Psi(N-D-n+1+\mu).$$

Let U be a uniform random variable. Accept the observation if

$U \leq R(\mu)$ or equivalently if $E = -\log U \geq -\log R(\mu)$ which reduces to $E + \log R(\mu) \geq 0$.

Chapter 4

Section 1: Simulation of the Model.

Let Ω be an environment. Recall that the species in the environment are to be such that for any interval $[a,b)$ with $0 < a < b < \infty$ the number of species with intensities in this interval has a Poisson distribution with mean $\int_a^b f(x)dx$ where $f(x) = A \frac{e^{-x}}{x^{\alpha}}$.

Suppose that A and α are given and that a sample of N individuals is to be taken from a computer simulated environment. The problem reduces to first choosing the intensities of the species in the environment so that they satisfy the above condition and then choosing the number of individuals in each species such that this number has a Poisson distribution with mean proportional to the intensity of the respective species. This constant of proportionality will be designated by k_s and will be called the power of the sample.

Let $x_1, x_2, \ldots$ be the intensities of the species that are to be selected and suppose a supply of independent exponential random variables $E_i, i = 1,2,\ldots$ are available.

Noting that the waiting time for a Poisson process is exponential consider the following method of choosing the intensities. Let

$$E_1 = \int_{x_1}^{\infty} f(x)dx \text{ and solve this equation for } x_1.$$

When this is done let

$$E_2 = \int_{x_2}^{x_1} f(x)dx \text{ and solve this equation for } x_2 \text{ and}$$

continue finding intensities $x_1, x_2, x_3, x_4, \dots$.

Notice that for $\epsilon > 0, \int_0^{\epsilon} f(x)dx = +\infty$ so that the method must be modified for small x. The modification and the method of determining a constant ϵ_s, which determines the intensity at which the modification will be made, will be shown later.

The function $f(x) = \frac{Ae^{-x}}{x^{\alpha}}$ cannot be integrated directly between two arbitrary positive numbers so that the solution of the equation $E_i = \int_{x_{i+1}}^{x_i} f(x)dx$ for x_{i+1} is obtained through an acceptance-rejection procedure. No such procedure was found that was effecient over the entire real line so that different procedures were used depending upon the portion of the real line that was being considered. The following method for finding the intensities of the species was used:

1. Set $y = x + \alpha \log x$

   so that $\frac{dy}{dx} = 1 + \frac{\alpha}{x} = \frac{x+\alpha}{x}$.

2. Let $y_0 = x_0 = +\infty$ and set $i = 1$, set $k = 1$ and

   set $E_0 = 0$.

3. Set $y_i^* = x_{i-1} + \alpha \log x_{i-1} = y_{i-1}$ and in order to determine x_i let

   $$y_i = x_i + \alpha \log x_i .$$

Then $\int_{x_i}^{x_{i-1}} f(x)dx = \int_{y_i}^{y_i^*} f(y) \frac{x}{x+\alpha} dy = \int_{y_i}^{y_i^*} Ae^{-y} \frac{x}{x+\alpha} dy$

4. Set $E_k = \int_{y_i}^{y_i^*} Ae^{-y}dy = A(e^{-y_i} - e^{-y_i^*})$

and solving for y_i

$$y_i = -\log (E_k + e^{-y_i^*}) + \log A$$
$$= -\log (E_k + \sum_{j=0}^{k-1} E_j) + \log A$$
$$= -\log (\sum_{j=1}^{k} E_j) + \log A.$$

5. Solve the equation $y_i = x_i + \alpha \log x_i$ for x_i.

6. Accept x_i with probability $\frac{x_i}{x_i+\alpha}$.

7a) If x_i is rejected set $y_i^* = y_i$, increase k by one and return to step #4 provided $x_i > 3.0$.

b) If x_i is accepted increase k by one, increase i by one and return to step #3 provided $x_i > 3.0$.

For intensities less than 3.0 a modification is made in the procedure to obtain a higher degree of effeciency in choosing the x_i.

8. Let x_i^* equal the last intensity calculated in step #5. Let $k_1 = k$, $x_{N_1} = x_i^*$.

Then $\int_{x_i}^{x_i^*} f(x)dx = \int_{x_i}^{x_i^*} \frac{Ae^{-x}}{x^\alpha} \frac{2^\alpha}{2^\alpha} dx = \int_{x_i}^{x_i^*} \frac{A}{2^\alpha} e^{-x} \left(\frac{2}{x}\right)^\alpha dx.$

9. Set $E_k = \int_{x_i}^{x_i^*} \frac{A}{2^\alpha} e^{-x}dx$ and solving for x_i

$$x_i = -\log \left[\frac{\sum_{j=k_1}^{k} E_j}{A\ 2^{-\alpha}} + e^{-x_{N_1}}\right].$$

10. If $x_i \geq 2.0$ accept x_1 with probability $\left(\frac{2}{x_i}\right)^\alpha$. If x_i is

rejected increase k by one, set $x_i^* = x_i$, return to step #9. If x_i is accepted increase k by one, set $x_{i+1}^* = x_i$, increase i by one, return to step #9.

For the case $x_i < 2.0$, x_i is rejected and the procedure again modified.

11. Let $x_i^* = 2.0$, increase k by one and set $k_2 = k$.

Then $\int_{x_i}^{x_i^*} f(x)dx = \int_{x_i}^{x_i^*} \frac{A\ e^{-x}}{x^\alpha} \frac{e^{-1}}{e^{-1}} dx = \int_{x_i}^{x_i^*} \frac{A\ e^{-1}}{x^\alpha} e^{1-x}\ dx.$

12. Set $E_k = \int_{x_i}^{x_i^*} \frac{A}{e} x^{-\alpha} dx$

and solving for x_i

$$x_i = \left[2^{1-\alpha} - \frac{(1-\alpha)e}{A} \left[\sum_{j=k_2}^{k} E_j\right]\right]^{\frac{1}{1-\alpha}} .$$

13. If $x_i \geq 1.0$ accept x_i with probability e^{1-x_i}. If x_i is rejected increase k by one, set $x_i^* = x_i$, return to step #12. If x_i is accepted increase k by one, set $x_{i+1}^* = x_i$, increase i by one, return to step #12. If $x_i < 1.0$, reject x_i since the procedure breaks down at one.

As was pointed out earlier, $\int_0^\epsilon f(x)dx = \infty$ for $\epsilon > 0$ so that procedures of the type used for large intensities are impractical for very small intensities. Note that when choosing a sample from the simulated environment the number of individuals in each species has a Poisson distribution with mean $k_s x_i$. Here k_s is unknown but it can be estimated and from this estimate a method devised for choosing species with small intensities which have a high probability of appearing in the sample while overlooking many which do not appear in the sample of N individuals.

The expected number of individuals is represented by the equation

$$\int_0^1 k_s x\, f(x)dx + \sum_{j=1}^{i-1} k_s x_j = \int_0^1 k_s x \frac{Ae^{-x}dx}{x^{\alpha}} + \sum_{j=1}^{i-1} k_s x_j$$

$$\approx k_s \left[\int_0^1 Ax^{1-\alpha}(1-x+\frac{x^2}{4})dx + \sum_{j=1}^{i-1} x_j\right] = k_s\left[A(\frac{1}{2-\alpha} - \frac{1}{3-\alpha} + \frac{1}{4(4-\alpha)}) + \sum_{j=1}^{i-1} x_j\right] .$$

Setting this equal to N, the number of individuals to be taken from the simulated environment, the estimate of k_s is

$$\hat{k}_s = \frac{N}{A(\frac{1}{2-\alpha} - \frac{1}{3-\alpha} + \frac{1}{4(4-\alpha)}) + \sum_{j=1}^{i-1} x_j}$$

Using this estimate of k_s continue finding the intensities of the species in the simulated environment.

14. Set $\epsilon_s = \frac{0.8}{\hat{k}_s}$, $x_i^* = 1.0$,

increase k by one and set $k_3 = k$. Then

$$\int_{x_i}^{x_i^*} f(x)dx = \int_{x_i}^{x_i^*} \frac{A}{x^{\alpha}} e^{-x}dx.$$

15. Set $E_k = \int_{x_i}^{x_i^*} \frac{A}{x^{\alpha}} dx$

and solving for x_i

$$x_i = \left[1 - \frac{(1-\alpha)}{A}\left[\sum_{j=k_3}^{k} E_j\right]\right]^{\frac{1}{1-\alpha}} .$$

16. Accept x_i with probability e^{-x_i}. If x_i is rejected, increase k by one, set $x_i^* = x_i$ and return to step #15 provided $x_i \geq \epsilon_s$. If x_i is accepted, increase k by one, set $x_{i+1}^* = x_i$, increase i by one and return to step #15 provided $x_i \geq \epsilon_s$.

For $x_i < \epsilon_s$ then the probability that a species with intensity x_i will have an individual present in the sample with power $\hat{k}_s$ is $1 - e^{-\hat{k}_s x_i}$. Therefore instead of solving the equation

$$E_k = \int_{x_i}^{x_i^*} f(x)dx$$

for x_i and letting the number of individuals present in the sample from this species be

$$\begin{cases} n_i & \text{where } n_i \sim \text{Truncated Poisson with parameter } \hat{k}_s x_i \text{ with probability } 1 - e^{-\hat{k}_s x_i} \\ 0 & \text{with probability } e^{-\hat{k}_s x_i} \end{cases}$$

an equivalent method for determining the individuals in the small species is to solve the equation

$E_k = \int_{x_i}^{x_i^*} \hat{k}_s x f(x)dx$ for x_i and let the number of individuals present in the sample from this species be

$$\begin{cases} n_i & \text{where } n_i \sim \text{Truncated Poisson with parameter } \hat{k}_s x_i \text{ with probability } \frac{1-e^{-\hat{k}_s x_i}}{\hat{k}_s x_i} \\ 0 & \text{with prob. } \frac{\hat{k}_s x_i - 1 + e^{-\hat{k}_s x_i}}{\hat{k}_s x_i} \end{cases}$$

This modification has the effect of skipping over some species which are in the environment but which do not appear in the sample.

17. Set $N^* = i$ and $x_\epsilon = x_i^*$, $k_4 = k$

$$\int_{x_i}^{x_i^*} \hat{k}_s x\, f(x)dx = \int_{x_i}^{x_i^*} \hat{k}_s x \frac{A}{x^\alpha} e^{-x}dx = \int_{x_i}^{x_i^*} \hat{k}_s A\, x^{1-\alpha} \quad e^{-x}dx.$$

18. Set $E_k = \int_{x_i}^{x_i^*} \hat{k}_s A \, x^{1-\alpha} dx$ and solving for x_i

$$x_i = \left(x_\epsilon^{2-\alpha} - \frac{(2-\alpha)\left[\sum_{j=k_4}^{k} E_j\right]}{\hat{k}_s A} \right)^{\frac{1}{2-\alpha}} .$$

19. If $x_i > 0$ accept x_i with probability e^{-x_1}.

If x_i is rejected, increase k by one, set $x_i^* = x_i$ and return to step #18. If x_i is accepted, increase k by one, set $x_{i+1}^* = x_i$, increase i by one and return to step #18.

The procedure is continued until a negative intensity is reached. Let s_N be the number of species obtained. Consider now the problem of finding the sample of N individuals and let n_i for $i = 1,2,\ldots,s_N$ be the number of individuals chosen from the species with intensity x_i. Thus

n_i is chosen from a Poisson distribution with parameter $\hat{k}_s x_i$ for $i = 1,2,\ldots,N^*-1$.

n_i chosen from a truncated Poisson distribution with parameter $k_s x_i$ with probability $\frac{1-e^{-k_s x_i}}{\hat{k}_s x_i}$,

0 with probability $\frac{\hat{k}_s x_i - 1 + e^{-\hat{k}_s x_i}}{\hat{k}_s x_i}$ for $i = N^*,\ldots,s_N$.

Let $N_T = \sum_{i=1}^{s_N} n_i$.

If $N_T = N$ then the sample is as chosen. If $N_T > N$ then $N_T - N$ individuals must be independently rejected from the chosen sample. This is accomplished be means of the Hypergeometric distribution where the number to be eliminated in the first species is n_1' which is distributed Hypergeometric with parameters $N_T, N_T - N, n_1$ and in general the number to be eliminated in the k^{th} species is

n'_k which is distributed Hypergeometric with parameter $N_T - \sum_{i=1}^{k-1} n_i$, $N_T - N - \sum_{i=1}^{k-1} n'_i$, n_k. This is continued until all N_T - N individuals have been eliminated.

The number of individuals in each species is

$n^*_i = n_i - n'_i$ for $i = 1,2,\ldots,s_N$ and $\sum_{i=1}^{s_n} n^*_i = N$.

If however $N_T < N$ then $N - N_T$ more individuals must be chosen from the model.

Let $\Delta\hat{k}_s = 2\hat{k}_s \dfrac{N-N_T}{N_T}$. The factor two is added to make the probability of falling short again vary small since it is better to over estimate k_s. The intensity of the sample is now $\hat{k}'_s = \hat{k}_s + \Delta\hat{k}_s$ so let n''_i be the number of individuals that are to be added to the already selected species where

$n''_i \sim$ Poisson $(\Delta\hat{k}_s x_i)$ for $i = 1,2,\ldots,s_N$.

Since some species were skipped in the interval $[0,x_\varepsilon)$ the possibility that some of these may now appear in the enlarged sample must be considered. Let $\varepsilon^* = \dfrac{1}{\hat{k}'_s}$. If $\varepsilon^* > x_\varepsilon$ select the new species using the method described in steps #17-19 replacing $\hat{k}_s$ by $\Delta\hat{k}_s$ and continue finding intensities until zero is reached. The number of individuals present in the sample from these species is n''_i where $n''_i \sim$ Truncated Poisson with parameter $\Delta\hat{k}_s x_i$ with

$$\text{probability } \frac{1 - e^{-\Delta\hat{k}_s x_i}}{\Delta\hat{k}_s x_i}$$

$$0 \quad \text{with probability } \frac{\Delta\hat{k}_s x_i - 1 + e^{-\Delta\hat{k}_s x_i}}{\Delta\hat{k}_s x_i}$$

for $i = s_N + 1,\ldots,s'_N$ where s'_N is now the number of species present. The intensity x_ε is chosen so that it is very unlikely that $\epsilon^* < x_\epsilon$ but if this should happen the situation can be corrected by decreasing the upper value of x_ϵ from say $\frac{0.8}{\hat{k}_s}$ to possibly $\frac{0.6}{\hat{k}_s}$ and rerunning the experiment.

Let $N'_T = \sum_{i=1}^{s'_N} n''_i$.

If $N'_T = N - N_T$ then no individuals need be deleted. If $N'_T > N-N_T$ then $N'_T - N + N_T$ individuals must be eliminated from the N'_T new individuals chosen.

This is accomplished again using the Hypergeometric distribution by letting n'_i be the number of individuals eliminated from the first species where n'_1 is distributed Hypergeometric with parameters N'_T, $N'_T - N + N_T$, n''_1 and for the k^{th} species n'_k is Hypergeometric $(N'_T - \sum_{i=1}^{k-1} n''_i, N'_T - N + N_T - \sum_{i=1}^{k-1} n'_1, n''_k)$.

The number of individuals in the respective species is then

$$n^*_i = n_i + n''_i - n'_i \qquad \text{for } i = 1,2,\ldots,s_N$$

$$n^*_i = n''_i - n'_i \qquad \text{for } i = s_N + 1,\ldots,s'_N$$

$$\text{where } \sum_{i=1}^{s'_N} n^*_i = N.$$

If $N'_T < N - N_T$ repeat the process for selecting new individuals from the species. Because of the method for determining $\Delta\hat{k}_s$ however it is extremely unlikely that the adding of new individuals will be necessary more than once.

Chapter 4

Section 2: Simulation in the Special case.

The simulation of the model for the special case developed in Chapter 1 Section 3 is greatly simplified over the general case. In taking the sample of size N in this case consider the following procedure. Define W_i as before to be the proportion of individuals of the i^{th} sampled species present in the environment neglecting the i-1 species already sampled.

Choose $w_1 \sim h(w) = A(1-w)^{A-1}$ where A is a parameter of the model which is to be estimated. Choose $m_1 \sim$ Binomial $(N-1, w_1)$ where m_1 represents the number of times that this species repeats in selecting the remaining N-1 individuals. Then $n_1 = m_1 + 1$ represents the number of individuals from the first species in the random sample of N individuals. Now choose $w_2 \sim h(w)$ and again select

$m_2 \sim$ Binomial $(N-n_1-1, w_2)$ and let $n_2 = m_2 + 1$.

In general select $w_i \sim h(w)$, select $m_i \sim$ Binomial $(N-\sum_{j=1}^{i-1} n_j - 1, w_i)$ and let $n_i = m_i + 1$.

Continue this process until a sample of N individuals has been chosen.

Chapter 5

Section 1: Data Analysis

Using the procedures developed in the previous chapters a set of data will now be analized to indicate the fit of the model in an actual environment. The data used for this purpose was taken from Williams [4] and is reproduced in table 5.1.

From the maximum likelihood methods developed in Chapter 2 an estimate of the parameters for this set of data was found to be

$\hat{\alpha} = 1.0000 \qquad \hat{A} = 40.453 \qquad \hat{k}_s = 392.7.$

Since $\hat{\alpha} = 1$ the estimation of the other two parameters reduces to the special case where α is set equal to one in the likelihood equations and an estimate of the other parameters obtained by the procedure developed in Chapter 2 Section 2. This maximum likelihood estimate was found to be

$$\hat{A} = 40.2576 \qquad \hat{k}_s = 387.2$$

According to the model each sample is such that the number of individuals in the sample from a species with intensity x is distributed Poisson with mean $k_s x$. Using $\hat{k}_s$ as an estimate of k_s then, the expected number of species in the sample with m individuals is

$$\int_0^\infty \frac{(\hat{k}_s x)^m}{m!} e^{-\hat{k}_s x} f(x)dx = \int_0^\infty \frac{(\hat{k}_s x)^m}{m!} e^{-\hat{k}_s x} \frac{\hat{A}e^{-x}}{x} dx$$

$$= \frac{\hat{A}\hat{k}_s^{\,m}}{m!} \int_0^\infty x^{m-1} e^{-(\hat{k}_s+1)x} dx = \frac{\hat{A}\,\hat{k}_s^{\,m}}{m!} \frac{\Gamma(m)}{(\hat{k}_s+1)^m}$$

$$= \hat{A}\left(\frac{\hat{k}_s}{\hat{k}_s+1}\right)^m \frac{1}{m} = \frac{40.2576}{m}(.99743)^m.$$

Table 5.1

Macrolepedoptera Data

Observed captures of Macrolepedoptera in a light trap at Rothamstad Journal of Animal Ecology Volume 12-13, pp.45-46.

Distribution of species according to number of individuals present in the sample.

	1	2	3	4	5	6	7	8	9	10
0	35	11	15	14	10	11	5	6	4	4
10	2	2	5	2	4	3	3	3	3	4
20	1	3	3	1	3	1	1	3	2	0
30	0	1	0	2	0	3	2	0	0	0
40	0	0	2	2	1	0	0	0	3	0
50	4	1	1	2	0	0	1	2	0	3

also at 61,64,67,73,76(2),78,84,89,96,99,109,112,120,122,129, 135,141,148,149,151,154,177,181,187,190,199,211,221,226,235,239, 244,246,282,305,306,333,464,560,572,589,604,743,823,2349

TOTAL NUMBER OF INDIVIDUALS 15,609

TOTAL NUMBER OF SPECIES 240

Table 5.2

Theoretical Frequencies for Macrolepedoptera Data

Distribution of the expected number of species present in the sample with parameters $\alpha = 1.0000$, $A = 40.2576$, $k_s = 387.2$

	1	2	3	4	5	6	7	8	9	10
0	40.15	20.03	13.31	9.96	7.96	6.61	5.64	4.93	4.37	3.92
10	3.55	3.25	2.99	2.77	2.58	2.41	2.26	2.13	2.02	1.91
20	1.81	1.73	1.65	1.58	1.51	1.45	1.39	1.34	1.29	1.24
30	1.20	1.16	1.12	1.08	1.05	1.02	.99	.96	.93	.91
40	.88	.86	.84	.81	.80	.78	.76	.74	.72	.71
50	.69	.68	.66	.65	.64	.62	.61	.60	.59	.57

also

61 - 70	5.14	151 - 200	7.31
71 - 85	6.34	201 - 300	8.57
86 -110	7.96	301 - 500	7.43
111 -150	8.84	500 ———	6.07

EXPECTED NUMBER OF INDIVIDUALS 15,587.7

EXPECTED NUMBER OF SPECIES 239.99

Table 5.3

Goodness of Fit Test

# of individuals in species	Observed frequency	Theoretical Frequency	$\frac{(f_{ob} - f_{th})^2}{f_{th}}$
1	35	40.15	.66
2	11	20.03	4.07
3	15	13.31	.21
4	14	9.96	1.64
5	10	7.95	.53
6	11	6.61	2.91
7	5	5.64	.07
8	6	4.93	.23
9-10	8	8.29	.01
11-12	4	6.80	1.15
13-14	7	5.77	.26
15-16	7	5.00	.80
17-19	9	6.40	1.06
20-22	8	5.44	1.20
23-25	7	4.71	1.11
26-30	7	6.68	.02
31-36	6	6.63	.06
37-45	7	7.98	.12
46-55	11	7.02	2.26
56-70	9	8.18	.08
71-85	5	6.34	.28
86-110	4	7.96	1.97
111-150	8	8.84	.08
151-200	7	7.31	.01
201-300	8	8.57	.04
301-500	4	7.43	1.58
500 —	7	6.07	.14
	240	239.99	22.55

$\chi^2_{.95}(24) = 36.42$

These values were calculated and are presented in table 5.2. It is interesting to note that in this special case the expected number of species present in the sample can be easly calculated by the formula

$$\int_0^\infty (1-e^{-\hat{k}_s x})\, f(x)dx = \int_0^\infty (1-e^{-\hat{k}_s x})\, \frac{\hat{A}e^{-x}}{x}\, dx$$

$$= \hat{A}\int_0^\infty \frac{e^{-x} - e^{-(\hat{k}_s+1)x}}{x}\, dx = \hat{A}\log(\hat{k}_s+1) = \hat{A}\cdot\hat{q} = 239.99.$$

Using these theoretical values a χ^2 goodness of fit test is applied to the data in table 5.1 and the theoretical values in table 5.2. The number of degrees of freedom for this test is j-3 where j is the number of categories. Here three degrees of freedom are lost because of the estimation of the three parameters of the model. The test is as shown in table 5.3 and is not significant at the 5% level.

As an aid in studying the behavior of the model a simulation procedure has been developed in the previous chapters. Three independent samples of 15,609 individuals have been taken from the model using the parameters estimated from the data in table 5.1 and the procedure developed in Chapter 4 Section 2. These three sets of simulated data are reproduced in tables 5.4-5.6 and should give the reader a good indication of the stability of the model. Note in particular that the total number of species present in each of the simulated samples are very close and that while the number of species present in the samples with a given number of individuals may have a large variation among samples nevertheless the number of large, moderate and small species

Table 5.4

Simulated Test #1

Distribution of species according to number of individuals present in the sample with parameters $\alpha = 1.0000$, $A = 40.2576$

	1	2	3	4	5	6	7	8	9	10
0	41	27	8	6	8	9	7	8	4	5
10	3	3	2	3	5	1	4	3	2	0
20	0	2	1	0	0	1	1	2	0	0
30	3	3	1	0	4	0	0	0	2	3
40	0	0	1	1	1	0	0	1	0	0
50	1	1	1	0	0	0	1	0	0	0

also at 62,63,66,67,79,80,83,85(2),88,89,91(2),93,94,96,97,105, 107,109(2),136,155,159(2),162,165,166,169,180,187,188,189,217, 222,246,247,255,260,273,277,287,324,325,345,350,405,408,440, 464,485,582,606,1385,1399.

TOTAL NUMBER OF INDIVIDUALS 15,609

NOTAL NUMBER OF SPECIES 235

Table 5.5

Simulated Test #2

Distribution of species according to number of individuals present in the sample with parameters α = 1.0000, A = 40.2576

	1	2	3	4	5	6	7	8	9	10
0	40	14	15	10	8	7	3	8	8	3
10	4	3	1	1	2	5	3	1	2	0
20	3	0	1	3	1	0	3	2	0	2
30	1	1	1	2	1	0	2	0	1	0
40	1	0	0	3	1	2	0	0	2	0
50	0	0	0	1	1	0	1	3	0	0

also at 61(4),64,66,67,69,71(2),72(3),73,74,75,93,94(2),97,100(2), 101,112,120,122,124(2),125,130,135,136,140,143,148,155,157,161, 175(3),177,187,191,192,193,196,205,206,237,291,295(2),299,302, 305,325,348,349,394,405,426,573,808,819,1079.

TOTAL NUMBER OF INDIVIDUALS 15,609

TOTAL NUMBER OF SPECIES 243

Table 5.6

Simulated Test #3

Distribution of species according to number of individuals present in the sample with parameters α = 1.0000 A = 40.2576

	1	2	3	4	5	6	7	8	9	10
0	45	18	12	6	10	7	4	8	3	4
10	6	3	1	2	1	1	2	3	2	2
20	3	0	1	1	2	2	2	0	1	0
30	2	1	3	1	4	1	3	0	1	0
40	0	0	1	0	2	0	0	0	0	1
50	4	0	2	0	0	1	1	1	1	0

also at 61,62,67,68,70(2),71,75,77,79,80(2),89,92(2),102,104, 105,106,107,113,115,119,121,125(2),138,152,168,192,196,208,218, 223(2),248,249,286,301,305,375,384,410,451,531,616,630,762,768, 1186,1711

TOTAL NUMBER OF INDIVIDUALS 15,609

TOTAL NUMBER OF SPECIES 233

present remains quite stable among samples.

With the use of the simulated tests the question of the accuracy of the estimates of the parameters when using the model can be considered. The simulated data is now considered as the original data to find the maximum likelihood estimates of the parameters, again using the procedures developed in Chapter 2. These estimates for the three simulated tests can be compared to the values of the parameters used in obtaining the simulated data as shown in the table below:

	α	A	k_s
Values of parameters	1.000	40.2576	387.2
Estimates for simulated test #1	1.000	39.2429	397.7
Estimates for simulated test #2	1.000	40.8523	382.1
Estimates for simulated test #3	1.000	38.8425	401.8

Another point of interest is to consider the behavior of the data as the number of individuals increases in the sample. Taking $\alpha = 1$ and $A = 40.2576$ table 5.7 shows the behavior of the data where a sample of size 50 is first taken and then the sample increased in small steps up to 15,609. It is to be remembered that this collection of data only illustrates the behavior as n increases in one sample but should serve as a guide for other samples. It is to be noted for example that the number of species with one individual in the sample has already stabilized by the time 200 individuals are sampled.

In order to compare the simulated data to the theoretical distribution for arbitrary N it is necessary to obtain an estimate of the parameter k_s. Noting that the number of individuals present in a sample from a species with intensity x_i is distributed Poisson with mean $k_s x_i$ so that the expected number is $k_s x_i$, an estimate of this parameter for arbitrary N is obtained by setting the equation $\int_0^\infty k_s x\, f(x)dx$ equal to N. Thus

$$\int_0^\infty k_s x \frac{Ae^{-x}}{x}\, dx = k_s A \int_0^\infty e^{-x} dx = k_s A$$

and the estimate is $\hat{k}_{s_N} = \frac{N}{A}$.

Using this estimate the expected number of species present in the sample with m individuals for a sample with parameter $\hat{k}_{s_N}$ and where $f(x) = \frac{Ae^{-x}}{x}$ is

$$\int_0^\infty \frac{(\hat{k}_{s_N} x)^m}{m!} e^{-\hat{k}_{s_N}} f(x)dx = \frac{\hat{k}^m_{s_N} A}{m!} \int_0^\infty x^{m-1} e^{-(\hat{k}_{s_N}+1)x} dx$$

$$= \frac{A\, \hat{k}^m_{s_N}\, \Gamma(m)}{m!\, (\hat{k}_{s_N}+1)^m} = \frac{A}{m} \left(\frac{\hat{k}_{s_N}}{\hat{k}_{s_N}+1} \right)^m$$

$$= \frac{A}{m} \left(\frac{N}{N+A} \right)^m .$$

For given values of N and A this can easily be tabulated and in particular compared with the data in table 5.7 for A = 40.2576.

Table 5.7

Distribution of species according to number of individuals present in the sample with parameters $\alpha = 1.0000$, $A = 40.2576$ for increasing N

N = 50 Number of species = 28

	1	2	3	4	5	6	7	8	9	10
0	15	8	2	2	1	0	0	0	0	0

N = 100 Number of species = 41

	1	2	3	4	5	6	7	8	9	10
0	20	8	3	2	5	1	0	2	0	0

N = 200 Number of species = 63

	1	2	3	4	5	6	7	8	9	10
0	29	10	8	2	3	2	2	3	1	0

also at 12(2),21

N = 500 Number of species = 94

	1	2	3	4	5	6	7	8	9	10
0	31	19	8	9	4	1	5	2	0	1
10	3	1	1	0	1	0	0	0	2	0
20	0	2	1	1	0	0	0	0	0	0

also at 33,49

N = 1000 Number of species = 119

	1	2	3	4	5	6	7	8	9	10
0	32	19	12	10	6	8	2	3	1	2
10	2	0	2	2	1	1	1	2	1	0
20	0	0	0	1	0	1	1	0	1	0

also at 33,42,43,46,47,49,67,97

N = 2000   Number of species = 143

	1	2	3	4	5	6	7	8	9	10
0	36	14	9	10	7	10	5	5	4	2
10	1	5	3	5	0	0	1	1	0	1
20	1	0	1	1	1	0	0	2	0	3

also at 34,35,40,47,49,54,56,59,80,83,93,95,97,144,203

N = 3000   Number of species = 165

	1	2	3	4	5	6	7	8	9	10
0	43	16	11	8	8	7	4	3	6	6
10	4	2	2	3	2	0	1	2	2	3
20	3	2	0	1	0	2	0	0	1	0

also at 39(2),41,42,44(2),47,50(2),52(2),67,79,80,81,92,124,
125,130,136,142,208,319

N = 4000   Number of species = 176

	1	2	3	4	5	6	7	8	9	10
0	41	22	5	11	5	5	7	6	3	4
10	4	2	4	5	1	4	1	2	2	1
20	2	2	1	1	0	1	2	0	1	0
30	2	2	0	2	0	0	0	1	1	0
40	0	0	0	0	0	0	0	0	1	2
50	0	0	1	0	0	0	1	0	0	0

also at 61,63,66,67,68,71,86,93,106,110,114,157,161,170,173,
190,295,434

N = 5000   Number of species = 190

	1	2	3	4	5	6	7	8	9	10
0	43	28	5	8	5	5	7	3	5	5
10	6	0	5	3	5	1	2	2	2	0
20	2	2	3	2	1	0	0	2	0	2
30	0	0	3	3	0	1	0	1	0	1
40	0	1	1	0	0	0	1	0	0	0
50	0	0	1	0	0	0	0	0	1	0

also at 61,64,65,71,74,79,82,88(2),89,98,118,130,133,138,195, 200,212,227,237,378,529

N = 6000   Number of species = 198

	1	2	3	4	5	6	7	8	9	10
0	40	31	8	7	7	5	6	0	5	5
10	2	5	5	2	1	4	7	1	1	0
20	1	0	3	2	3	1	0	1	1	1
30	1	1	1	1	1	1	1	0	1	1
40	0	2	3	0	1	0	1	0	0	0
50	0	0	1	2	0	1	0	0	0	0

also at 70(2),77,81,87,91,93,96,101,107,120(2),142,149,150, 171,218,236,248,275,285,449,633

N = 7000   Number of species = 200

	1	2	3	4	5	6	7	8	9	10
0	35	29	9	8	11	2	7	4	1	5
10	4	1	4	6	2	3	2	1	1	5
20	0	1	1	1	2	2	2	2	1	0
30	1	0	2	1	1	2	1	2	0	0
40	0	2	0	0	1	0	0	2	1	0
50	3	1	0	0	0	1	0	1	0	1

also at 61,67,81,83,86,92,102,105,109(2),115,124,136,141,172, 173,175,197,256,277,288,332(2),531,742

N = 8000 Number of species = 205

	1	2	3	4	5	6	7	8	9	10
0	35	23	16	7	14	3	3	3	5	1
10	4	1	5	1	6	3	3	1	1	2
20	0	3	2	1	2	2	3	0	3	0
30	1	1	2	1	0	1	1	0	2	1
40	0	2	0	1	0	0	1	2	0	1
50	0	0	1	3	0	1	0	0	0	1

also at 61,64,66,67,72,75,94,95,97,108,111,116,122,123,129,149, 150,167,192,195,204,227,284,314,325,382,393,600,860

N = 9000 Number of species = 207

	1	2	3	4	5	6	7	8	9	10
0	35	22	13	8	9	9	0	7	2	4
10	5	2	2	1	6	3	3	1	1	3
20	1	0	1	2	2	2	3	1	2	1
30	1	1	0	2	0	2	1	0	1	1
40	0	1	1	1	2	1	1	0	0	1
50	0	0	0	1	1	0	1	0	3	0

also at 61,64,65(2),68,71,72,77,81,84,99,105(2),118,120,121, 131,140,147,168,173,182,213,224,234,265,324,360,365,433,445, 676,961

N = 10,000 Number of species = 210

	1	2	3	4	5	6	7	8	9	10
0	32	25	13	7	9	6	4	6	5	4
10	1	3	3	3	2	2	2	3	3	2
20	2	2	0	2	1	0	1	1	1	4
30	3	0	1	0	1	4	0	0	1	0
40	0	1	1	2	0	1	0	3	1	0
50	1	1	1	0	0	0	0	0	1	0

also at 61,62,65(2),71(2),72(2),74(2),76,80,86,88,93,112,113, 123,130,132,138,141,154,160,191,192,199,236,244,274,293,355,391, 397,479,492,751,1093

N = 11,000 Number of species = 215

	1	2	3	4	5	6	7	8	9	10
0	34	25	12	8	9	6	4	7	4	2
10	4	3	0	4	1	3	4	3	1	0
20	5	1	1	1	1	3	1	1	0	1
30	1	1	1	2	2	1	1	3	0	1
40	1	0	0	0	2	1	1	0	0	1
50	1	1	2	0	1	1	0	0	2	0

also at 64,66,69,70,71,77(3),79,80,81,85(2),95,96,105,121,126, 135,145,148,154,155,173,175,207,213(2),256,261,298,316,385,440, 442,529,533,835,1215

N = 12,000 Number of species = 220

	1	2	3	4	5	6	7	8	9	10
0	37	23	12	8	7	11	2	6	5	2
10	3	4	2	2	1	2	2	6	1	1
20	2	1	1	2	2	2	2	2	2	0
30	1	0	1	0	1	2	2	1	0	3
40	0	1	1	1	0	1	0	1	2	0
50	0	2	0	1	0	1	1	1	0	2

also at 62,63,67,70,73,80,81,85(2),86(2),88,89,93,94,101,105, 121,128,144,148,157,162,169,171,187,189,223,227,230,276,279,322, 348,423,474,484,575,592,916,1327

N = 13,000 Number of species = 221

	1	2	3	4	5	6	7	8	9	10
0	38	20	10	11	7	9	5	1	9	2
10	1	2	5	4	1	2	2	2	2	3
20	1	1	0	1	3	1	3	1	3	2
30	1	1	1	1	0	0	1	0	3	1
40	0	1	1	0	2	2	0	1	1	0
50	1	1	0	1	1	0	0	1	2	1

also at 64(2),65,66,67,71,74,77,84,90,92(2),94,96,98,99,101, 102,106,114,130,138,161,165,177,181,189,201,207,239,247,252, 305,311,342,374,452,515,518,628,646,989,1415

N = 14,000 Number of species = 224

	1	2	3	4	5	6	7	8	9	10
0	39	20	10	10	8	8	5	2	8	5
10	1	2	1	3	2	3	4	1	2	3
20	1	1	2	0	1	1	1	2	2	0
30	1	5	4	0	0	1	0	0	1	0
40	0	1	3	0	0	1	2	1	1	0
50	1	1	1	1	0	0	1	1	1	0

also at 62,63(2),64,69,70,71,72(2),78,84(2),90,96,97(2),99, 103,106(2),107,109,116,123,136,148,171,179,194,195,197,203,221, 223,258,269(2),333,339,400,477,556,559,679,684,1063,1528

N = 15,000 Number of species = 230

	1	2	3	4	5	6	7	8	9	10
0	43	17	13	6	10	9	5	4	4	5
10	5	2	1	2	3	0	2	3	3	2
20	2	1	2	1	3	0	0	0	1	1
30	2	2	2	3	5	0	0	1	0	0
40	0	1	0	1	1	1	0	0	3	2
50	0	1	1	0	0	1	2	0	0	1

also at 62,64,65,66,68(2),73,75(2),77,79,83,87(2),99,101,102, 103,106,108,111,115,117,119,123,133,144,164,184,190,205,206,213, 216,235,238,275,287,294,356,369,399,430,508,588,602,728,741,1134, 1647

N = 15,609 Number of species = 233

	1	2	3	4	5	6	7	8	9	10
0	45	18	12	6	10	7	4	8	3	4
10	6	3	1	2	1	1	2	3	2	2
20	3	0	1	1	2	2	2	0	1	0
30	2	1	3	1	4	1	3	0	1	0
40	0	0	1	0	2	0	0	0	0	1
50	4	0	2	0	0	1	1	1	1	0

also at 61,62,67,68,70(2),71,75,77,79,80(2),89,92(2), 102,104, 105,106,107,113,115,119,121,125(2),138,152,168,192,196,208,218, 223(2),248,249,286,301,305,375,384,410,451,531,616,630,762,768, 1186,1711.

Chapter 6

Section 1: Investigation of Species per Genus Data

In an effort to determine the different types of environments for which the model holds data from Williams[6] on Orthoptera was investigated. It is realized that the data is in the form of species per genus which is quite a different concept from the individuals per species data that had previously been considered but this data seemed to show some of the same properties as the other data and it was hoped that this biological situation could also be explained by the model. Applying therefore the methods of the previous chapters the maximum likelihood estimate of the parameters was

$$\hat{\alpha}=1.1056 \qquad \hat{A}=231.065 \qquad \hat{k}_s=16.3$$

In comparing the actual data , reproduced in table 6.1, to the theoretical expected values obtained using the above estimates of the parameters it was determined that the model fit rather well for the small and moderate genera but that the theoretical values for the larger genera were too small. This conclusion was reinforced when three samples of 4112 species were taken using $\alpha=1.1056$ and $A=231.065$ and it was found that the largest genus among the three samples contained only 80 species, far below the number that was actually encountered.

With this result in mind it was decided that an adequate fit might be obtained if the form of the function f(x) was altered to accommodate this new situation . It was decided that the term e^{-x} in the numerator made the function f(x) decrease too rapidly. For large intensities it was decided to try the form $f(x)=\frac{A}{x^q}$ where q is a parameter with the restriction $2\leq q<\infty$.

In determining q for the Orthoptera data make the definition

SPC[a,b) = total number of species in the genera which have n_i species with $a \le n_i < b$. Adjusting so that $k_s = 1$ the expected number of species is defined by the equation $\int_a^b x\, f(x)\, dx$ for the interval [a,b).

Using SPC[a,b) as an estimate of the number of species in the sample which are in genera having an intensity in the interval [a,b) consider the following equations

$$\int_{30}^{c} x\, f(x)\, dx = \int_{30}^{c} A\, x^{1-q}\, dx = SPC[30,c)$$

$$\int_{c}^{\infty} x\, f(x)\, dx = \int_{c}^{\infty} A\, x^{1-q}\, dx = SPC[c,\infty)$$

for $50 < c < 100$
and $q > 2$

Integrating and eliminating A from the two above equations the solution for q is seen to be

$$q = 2 - \frac{\log SPC[30,\infty) - \log SPC[c,\infty)}{\log 30 - \log c}$$

From the graph of q as a function of c for $50 < c < 100$ a good choice for q in this case seems to be q = 3. Also for small genera the function f(x) appears to take the general form similar to $f(x) = \frac{A}{x}$ so that the expected number of genera with m species is

$$\int_0^{\infty} \frac{x^m e^{-x}}{m!} f(x)\, dx = \int_0^{\infty} \frac{A\, x^{m-1} e^{-x}}{m!}\, dx = \frac{A\, \Gamma(m)}{m!} = \frac{A}{m}$$

Combining these two characteristics it was decided that the function f(x) should take the form $f(x) = \frac{A}{x(x+a)^2}$ where A and a are positive constants. An attempt at finding a maximum likelihood estimate of these parameters became very messy so that an estimate was obtained from a simultaneous solution of the equations

$$\int_{30}^{\infty} x\, f(x)\, dx = \int_{30}^{\infty} \frac{A}{(x+a)^2}\, dx = SPC[30,\infty) = 904$$

$$\int_{0}^{\infty} \frac{x\, e^{-x}}{1!}\, f(x)\, dx = \int_{0}^{\infty} \frac{A\, e^{-x}}{(x+a)^2}\, dx = 320\ .$$

The estimates obtained were

$$\hat{a} = 10 \qquad \hat{A} = 37{,}000$$

The simulation procedure used in the case where $f(x) = \frac{A}{x(x+a)^2}$ is quite similar to the procedure developed in Chapter 4 Section 1 except that some changes are needed in finding the intensities due to the different form of the function f(x). As before let E_k, k=1,2,3,... be an infinite supply of exponential random variables and consider the following procedure for producing a sample of size N with known parameters a and A.

1. Set k=1, set i=1, set $x_i^* = +\infty$

2. $$\int_{x_i}^{x_i^*} f(x)\, dx = \int_{x_i}^{x_i^*} \frac{A}{x(x+a)^2}\, dx = \int_{x_i}^{x_i^*} \frac{A}{x^3}\left(\frac{x}{x+a}\right)^2 dx$$

3. Set $E_k = \int_{x_i}^{x_i^*} \frac{A}{x^3}\, dx$ and solving for x_i

$$x_i = \sqrt{\frac{A}{2}\, \frac{1}{\sum_{j=1}^{k} E_j}}$$

4. Accept x_i with probability $\left(\frac{x_i}{x_i + a}\right)^2$.

5. If x_i is rejected, set $x_i^* = x_i$, increase k by one, and return to step #3 provided $x_i > \frac{a}{\sqrt{2} - 1}.$ If x_i is accepted, increase k by one, set $x_{i+1}^* = x_i$, increase i by one and return to step #3 provided $x_i > \frac{a}{\sqrt{2} - 1}$.

The point $x_i = \frac{a}{\sqrt{2} - 1}$ is the point where the acceptance probability $\left(\frac{x}{x+a}\right)^2$ is equal to one half so that it becomes desirable to modify the procedure at this point to increase the effeciency.

6. Set $k_1 = k$. Also

$$\int_{x_i}^{x_i^*} f(x)\,dx = \int_{x_i}^{x_i^*} \frac{A}{x(x+a)^2}dx = \int_{x_i}^{x_i^*} \frac{A}{2x^3}\,\frac{2}{1}\left(\frac{x}{x+a}\right)^2 dx$$

7. Set $E_k = \int_{x_i}^{x_i^*} \frac{A}{2}\,\frac{1}{x^3}dx$ and solving for x_i

$$x_i = \sqrt{\frac{A}{2}}\;\frac{1}{\sqrt{2\sum_{j=k_1}^{k} E_j + \sum_{j=1}^{k_1-1} E_j}}$$

8. Accept x_i with probability $\frac{2x_i^2}{(x_i+a)^2}$.

9. If x_i is rejected, set $x_i^* = x_i$, increase k by one, and return to step #7 provided $x_i > a$. If x_i is accepted, set $x_{i+1}^* = x_i$, increase k by one, increase i by one and return to step #7 provided $x_i > a$.

At the point $x = a$ another modification is to be made to increase the effeciency.

10. Set $k_2 = k$, set $x_{N_1} = x_i^*$

$$\int_{x_i}^{x_i^*} f(x)\,dx = \int_{x_i}^{x_i^*} \frac{A}{x(x+a)^2}dx = \int_{x_i}^{x_i^*} \frac{A}{xa^2}\left(\frac{a}{x+a}\right)^2 dx$$

11. Set $E_k = \int_{x_i}^{x_i^*} \frac{A}{xa^2}dx$ and solving for x_i

$$x_i = x_{N_1}\, e^{-\frac{a^2}{A}\left[\sum_{j=k_1}^{k} E_j\right]}.$$

12. Accept x_i with probability $\left(\frac{a}{x_i+a}\right)^2$.

13. If x_i is rejected, set $x_i^* = x_i$, increase k by one, and return to step #11 provided $x_i > \epsilon_s$. If x_i is accepted, set $x_i^*+1 = x_i$, increase k by one, increase i by one and return to step #11 provided $x_i > \epsilon_s$.

The constant ε_s is determined similar to the procedure used before.

The expected sample size is

$$\int_0^\infty k_s x\, f(x)dx = \int_0^\infty k_s x \frac{A}{x(x+a)^2} dx = k_s A \int_0^\infty \frac{1}{(x+a)^2}\, dx = \frac{k_s A}{a}.$$

Setting this equal to N to obtain an estimate for k_s

$$\hat{k}_s = \frac{aN}{A}.$$

14. Set $N^* = i$ and $k_4 = k$, $x_\varepsilon = x_i^*$.

For the small intensities the modification which skips over some of the genera which do not appear in the sample is again employed.

15. $$\int_{x_i}^{x_i^*} \hat{k}_s x\, f(x)dx = \int_{x_i}^{x_i^*} \hat{k}_s x \frac{A}{x(x+a)^2} dx = \int_{x_i}^{x_i^*} \frac{\hat{k}_s A}{a^2} \left(\frac{a}{x+a}\right)^2 dx.$$

16. Set $E_k = \int_{x_i}^{x_i^*} \frac{\hat{k}_s A}{a^2}\, dx$ and solve for x_i

$$x_i = x_\varepsilon - \frac{a^2}{\hat{k}_s A} \sum_{j=k_4}^{k} E_j.$$

17. If $x_i > 0$, accept x_i with probability $\left(\frac{a}{x_i+a}\right)^2$. If x_i is rejected increase k by one, set $x_i^* = x_i$ and return to step #16. If x_i is accepted, increase k by one, set $x_{i+1}^* = x_i$, increase i by one and return to step #16.

If $x_i \leq 0$, reject x_i and cease finding intensities.

In finding the number of species in each genus for a particular sample employ the procedure described in Chapter 4.

Using the above prodecure with A = 37,000 and a = 320 three samples of 4112 genera were taken and the results shown in table 6.2. These results can be compared to the original data in table 6.1 to examine the fit of the model in this case.

Table 6.1

ORTHOPTERA OF WORLD

Journal of Ecology Volume 32 page 18

Distribution of genera according to number of species present

	1	2	3	4	5	6	7	8	9	10
0	320	131	86	61	41	27	21	18	23	17
10	12	8	9	3	5	4	3	6	2	3
20	1	1	2	1	0	2	0	0	4	0

also at 31(2),34,35,36,38,41,43,51,54,58,72,75,103,202.

TOTAL GENERA 826

TOTAL SPECIES 4112

Table 6.2

SIMULATED TEST 1

Distribution of genera according to number of species present

	1	2	3	4	5	6	7	8	9	10
0	317	134	86	49	37	24	26	22	7	12
10	10	8	8	7	1	5	3	3	0	0
20	1	5	4	1	5	2	1	1	1	0

also at 32,34,35,36,44(2),49,53,54,55,56(2),57,69,73,79,83,178.

TOTAL GENERA 798

TOTAL SPECIES 4112

Table 6.2

SIMULATED TEST 2

Distribution of genera according to number of species present

	1	2	3	4	5	6	7	8	9	10
0	324	146	90	54	43	24	17	23	15	5
10	7	9	7	8	8	8	5	3	6	1
20	3	2	3	2	2	3	1	3	0	1

also at 32,34,35,36,39(2),41,45,49,52,53,74,79,153.

TOTAL GENERA 837

TOTAL SPECIES 4112

SIMULATED TEST 3

Distribution of genera according to number of species present

	1	2	3	4	5	6	7	8	9	10
0	317	141	83	48	41	25	21	9	15	7
10	7	9	7	4	5	2	4	4	4	5
20	3	5	1	3	2	0	1	1	2	2

also at 33,37,42,46,48,50(2),56,57(2),217,354.

TOTAL GENERA 790

TOTAL SPECIES 4112

APPENDIX

Using the theory developed in the previous chapters, FORTRAN 60 programs have been developed to perform the indicated operations on the Control Data 3600 computer.

Program SPECIES 1 finds the maximum likelihood estimates of the parameters of the model using the methods discribed in Chapter 2 Section 1.

Program SPECIES 2 finds the maximum likelihood estimates of the parameters of the model under the special condition $\alpha = 1$ using the methods discribed in Chapter 2 Section 1.

Program SPECIES 3 is a simulation program to obtain a sample of size N from the model in the case $\alpha \neq 1$ using the methods developed in Chapter 4 Section 1.

When using the program to obtain a sample it was found that about 5000 individuals could be sampled in about 30 seconds on the CDC 3600 computer. Also note that if the sample size is doubled the estimated simulation time increases only a few seconds due to the fact that a large percentage of the simulation time is used to obtain the intensities of the species and the number of new species decreases rapidly with increasing sample size.

Program SPECIES 4 is a simulation program to obtain a sample of size N from the model in the case $\alpha = 1$ using the procedure described in Chapter 4 Section 2.

This program obtains a sample of 15,000 individuals in about 20 seconds on the CDC 3600 computer. Note that this procedure is much faster than SPECIES 3. This is explained by the fact that the simulation procedure is extremely simplified in the case where $\alpha = 1$.

The four above mentioned programs are tabulated in the following pages with a brief explination to the right of the tabulated programs. Although these were not the only programs used in this investigation, they were the ones used to obtain the primary results.

```
   PROGRAM SPECIES 1
 1 DIMENSION SOLVE(3),EST(3),COF(12),IDATA(200)
 4 FORMAT(3(2X,I3))
 5 FORMAT(14(1X,I4))
 6 FORMAT(3E20.11)
10 READ INPUT TAPE2,4,K1,K2,K3                          } INPUT DATA
12 READ INPUT TAPE2,5,(IDATA(I),I=1,K1)
13 READ INPUT TAPE2,6,(COF(I),I=1,12)
   KJ=0
89 M=0
   READ INPUT TAPE 2,6,(EST(I),I=1,3)                   INITIAL ESTIMATE OF PARAMETERS
50 SOLV8=-EST(2)
51 SOLV6=-(EST(3)-1.0)*SOLV8/2.0
52 IF(1.0-SOLV6) 53,53,55
53 ETA1=(1.0-EXPF((EST(3)-1.0)*SOLV8))/(EST(3)-1.0)
54 GOTO 58
55 SOLV7=SOLV6**2
56 Q1=((105.0+SOLV7)*SOLV7+945.0)/((420.0+15.0*SOLV7)*SOLV7+945.0)
57 ETA1=-Q1*SOLV8/(1.0+Q1*SOLV6/2.0)**2
58 IF(1.5-EST(3))59,59,63
59 X=2.0-EST(3)
60 CALL GAMMA(X,GAM2)
61 GAM1=GAM2/(1.0-EST(3))                               CALCULATE Γ(1-α), Γ(2-α)
62 GOTO 66
63 X=1.0-EST(3)
```

```
64 CALL GAMMA(X,GAM1)
65 GAM2=GAM1*(1.0-EST(3))
66 X=12.0-EST(3)
67 CALL PSI(X,PSI1,PSI2)
   PSIM=PSI1
   PSIN=PSI2
   SUM1=0.0
   SUM2=0.0
   ISPEC=0
   INDIV=0
   DO 70 I=1,12
   ISPEC=ISPEC+IDATA(13-I)
   DATA=IDATA(13-I)
   SUM1=SUM1+DATA*PSIM
   SUM2=SUM2+DATA*PSIN
   INDIV=INDIV+(13-I)*IDATA(13-I)
   Y=I-1
   PSIM=PSIM-1.0/(X-Y)
70 PSIN=PSIN+1.0/(X-Y)**2
   DO 71 I=13,K2
   ISPEC=ISPEC+IDATA(I)
   INDIV=INDIV+I*IDATA(I)
   DATA=IDATA(I)
   Y=I-12
   PSI1=PSI1+1.0/(X+Y)
```

$$SUM\ 1 = \sum_{n=1}^{\infty} x_n\, \psi(n-\alpha+1)$$

$$SUM\ 2 = \sum_{n=1}^{\infty} x_n\, \psi'(n-\alpha+1)$$

rSIM - ♦(1-ø)

```
      PSI2=PSI2-1.0/(X+Y)**2
      SUM2=SUM2+DATA*PSI2
   71 SUM1=SUM1+DATA*PSI1
      K4=K2+1
      DO 72 I=K4,K3
      X=IDATA(I)
      X=X-EST(3)
      CALL PSI(X,PSI1,PSI2)
      ISPEC=ISPEC+1
      INDIV=INDIV+IDATA(I)
      SUM1=SUM1+PSI1
   72 SUM2=SUM2+PSI2
      DATA=EXPF(SOLV8*(EST(3)-1.0))
      SOLV1=ISPEC
 1000 EST(1)=SOLV1/(GAM2*ETA1)
      SOLV1=SOLV1/EST(1)
      SOLV2=-GAM2*DATA
      SOLV7=GAM2*ETA1
      SOLV3=SOLV7*PSIM
      SOLV6=GAM1*SOLV8*DATA
      SOLV3=SOLV3-SOLV6
      SOLV5=INDIV
      SOLV4=-EST(1)*GAM2*DATA*(1.0-EST(3))-SOLV5*EXPF(EST(2))/
     1(EXPF(EST(2))-1.0)**2
 1001 SOLVE(1)=0.0
```

PSIM = $\psi(1-\alpha)$

PSIN = $\psi'(1-\alpha)$

NEW ESTIMATE FOR A

EVALUATING EQUATIONS FOR NEWTON PROCEDURE

```
      SOLV1=-SOLV1/EST(1)
      SOLVE(2)=-EST(1)*SOLV2-SOLV5/(EXPF(EST(2))-1.0)
      SOLV5=EST(1)*(GAM1*DATA              +SOLV2*(SOLV8-PSIM))
      SOLVE(3)=-EST(1)*SOLV3+SUM1
      SOLV6=-EST(1)*(SOLV3*PSIM+SOLV7*PSIN+SOLV6*(SOLV8-PSIM))+SUM2
      KJ=1-KJ
      IF(KJ-1) 120,100,120
  100 SOLVE(2)=SOLV1*SOLVE(2)/(SOLV1*SOLV4-SOLV2**2)
      EST(2)=EST(2)+SOLVE(2)                         NEW ESTIMATE FOR q
      GOTO 156
  120 SOLVE(3)=SOLV1*SOLVE(3)/(SOLV1*SOLV6-SOLV3**2)
  147 ESTA=(EST(3)-1.0)/2.0                          }
  148 IF(SOLVE(3)+ESTA) 149,149,157                  }
  149 EST(3)=1.0+ESTA                                }
  150 GOTO 152                                       }
  157 ESTA=(2.0-EST(3))/2.0                          } NEW ESTIMATE FOR α
      IF(ESTA-SOLVE(3)) 158,158,151                  }
  158 EST(3)=EST(3)+ESTA                             }
      GOTO 152                                       }
  151 EST(3)=EST(3)+SOLVE(3)                         }
  152 CONTINUE
      WRITE OUTPUT TAPE3,6,(EST(I),I=1,3)            PRINT NEW ESTIMATES
  160 M=M+1                                          STOP AFTER 50 ITERATIONS
  161 IF (M-50)  50,180,180
  180 STOP 0001
```

```
      END
      SUBROUTINE PSI(X,PSI1,PSI2)
      X2=X**2
      PSI1=((((0.0125/X2-1.0/84.0)/X2+0.025)/X2-0.25)/(3.0*X)+0.5)/X+
     1LOGF(X)
      FSI2=(((((-0.1/X2+1.0/14.0)/X2-0.1)/X2+0.5)/(3.0*X)-0.5)/X+1.0)/X
      RETURN
      END
      SUBROUTINE GAMMA(X,GAMMAT)
      DIMENSION COF(12)
      GAMMAT=0.0
      DO 75 I=1,12
   75 GAMMAT=(GAMMAT+COF(13-I))*X
      GAMMAT=1.0/GAMMAT
      RETURN
      END
      END
```

(PSI1 through FSI2: CALCULATE ψ(x) and ψ'(x))

(GAMMAT=0.0 through RETURN: CALCULATE Γ(X))

INPUT DATA

+0.10000000000E 01	+0.57721566490E 00	-0.65587807152E 00
-0.42002635034E-01	+0.16653861138E 00	-0.42197734556E-01
-0.96219715279E-02	+0.72189432467E-02	-0.11651675919E-02
-0.21524167411E-03	+0.12805028439E-03	-0.20134854781E-04

(COF MATRIX)

ESTIMATES OF PARAMETERS

```
      PROGRAM SPECIES 2
    1 DIMENSION IDATA(200)
    2 FORMAT(2E20.11)
    4 FORMAT(3(2X,I3))
    5 FORMAT(14(1X,I4))
   10 READ INPUT TAPE2,4,K1,K2,K3                          } INPUT DATA
   12 READ INPUT TAPE2,5,(IDATA(I),I=1,K1)
      M=0
      ISPEC=0
      INDIV=0
      DO 30 I=1,K2
      ISPEC=ISPEC+IDATA(I)                                 ISPEC = NUMBER OF SPECIES
   30 INDIV=INDIV+I*IDATA(I)                               INDIV = NUMBER OF INDIVIDUALS
      K4=K2+1
      DO 31 I=K4,K3
      ISPEC=ISPEC+1
   31 INDIV=INDIV+IDATA(I)
      SPEC=ISPEC
      SINDIV=INDIV
      BAMDA=SINDIV/SPEC
      QQ=LOGF(1.0+BAMDA*LOGF(BAMDA))                       } INITIAL ESTIMATE OF q and e^q
   32 EQ=EXPF(QQ)
      AQ=EQ-1.0-BAMDA*QQ
      AA=(EQ-AQ/18.0)/2.0                                  SOLVE QUADRATIC
      AB=EQ-BAMDA+AQ/3.0
```

NEW ESTIMATE

```
      X=-2.0*AQ/(AB+SQRTF(AB**2-4.0*AA*AQ))
      QQ1=QQ+X/(1.0+X/3.0-X**2/36.0)          NEW ESTIMATE OF q
      IF(ABSF((QQ-QQ1)/QQ1)-0.00001) 33,33,34  PRINT WHEN FIVE DIGIT ACCUEACY IS
                                               ATTAINED
   33 BETA=SPEC/QQ1
      WRITE OUTPUT TAPE3,2,QQ1,BETA            PRINT ESTIMATES
      STOP 0002
   34 QQ=QQ1
      M=M+1
      IF(M-10) 32,32,35
   35 STOP 0001                                STOP PROCEDURE AFTER TEN
                                               ITERATIONS.
      END
      END
```

```
      PROGRAM SPECIES 3
      DIMENSION CLASS(2000),INCLS(2000),PSIII(11),RANDOM(72),INCL2(2000)
 1481 FORMAT(10(I5))
 1482 FORMAT(2X,I5)
 1485 FORMAT(I5,3(E15.5))
      PAUSE 0001
 1200 X=1.0/10.5**2
 1201 PSIII(11)=((((+0.0008401068*X-0.00059058779)*X+0.0007688492)*X      }
     1-0.0024305555)*X+0.0416666666)/10.5                                  }
 1202 DO 1206 J=1,10                                                       }
 1203 PNI=11-J                                                             }  CALCULATE  Ψ(I) I = 1,2,..11
 1204 PNI=1.0/(2.0*PNI)**2                                                 }
 1205 DELTA=((((((PNI/210.0+1.0/156.0)*PNI+1.0/110.0)*PNI+1.0/72.0)*       }
     1PNI+1.0/42.0)*PNI+0.05)*PNI+1.0/6.0)*PNI                             }
 1206 PSIII(11-J)=PSIII(12-J)-DELTA                                        }
 1207 PSIII(1)=PSIII(1)-(((0.25/506.0+1.0/421.0)*0.25+1.0/342.0)*0.25      }
     1+1.0/272.0)*0.25**8
 1210 CALL RANREAD(RANDOM,KZ)
      READ INPUT TAPE2,1485,ISAMP,ALPHA,BETA,QQQ                              INPUT DATA
 1296 BETA=BETA*EXPF(-QQQ*(ALPHA-1.0))
 1297 ALPHA2=2.0**(-ALPHA)
 1298 BETALG=LOGF(BETA)
 1299 SUM =0.0
 1300 I=0
 1302 I=I+1
```

```
1303 CALL EXPVAR2(X,RANDOM,KZ)
1304 SUM =SUM+X
1305 CLASSS  =-LOGF(SUM )+BETALG
1307 CALL FITG(CLASSS  ,ALPHA,CLASS(I))
1308 TEST=CLASS(I)/(CLASS(I)+ALPHA)
1309 CALL UNIFORM(X,RANDOM,KZ)
1310 IF(TEST-X) 1303,1311,1311
1311 IF(CLASS(I)-3.0) 1312,1302,1302
1312 SUM =0.0
     SUM1=EXPF(-CLASS(I))
```

FINDING INTENSITIES $X_i \geq 3.0$

```
1314 I=I+1
1315 CALL EXPVAR2(X,RANDOM,KZ)
1316 SUM =SUM+X
1317 TOTL=SUM /(ALPHA2*BETA)+SUM1
1318 CLASS(I)=-LOGF(TOTL)
1319 TEST=(2.0/CLASS(I))**ALPHA
1320 CALL UNIFORM(X,RANDOM,KZ)
1321 IF(TEST-X) 1315,1322,1322
1322 IF(CLASS(I)-2.0) 1324,1314,1314
```

FINDING INTENSITIES $2.0 \leq X_i < 3.0$

```
1324 SUM =0.0
1325 SUM3=2.0**(1.0-ALPHA)
1327 CALL EXPVAR2(X,RANDOM,KZ)
1328 SUM =SUM+X
1329 CLASS(I)=(SUM3-SUM *(1.0-ALPHA)*2.7182818284/BETA)**(1.0/(1.0-
    1ALPHA))
```

FINDING INTENSITIES $1.0 \leq X_i < 2.0$

```
1330 TEST=EXPF(1.0-CLASS(I))
1331 CALL UNIFORM(X,RANDOM,KZ)
1332 IF(TEST-X) 1327,1333,1333
1333 IF(CLASS(I)-1.0) 1336,1334,1334
1334 I=I+1
1335 GOTO 1327
1336 TOTL=0.0
     KK=I-1
1337 DO 1338 J=1,KK
1338 TOTL=TOTL+CLASS(J)
1339 TOTL=TOTL+BETA*(1.0/(2.0-ALPHA)-1.0/(3.0-ALPHA)+0.25/(4.0-ALPHA))
     SAMPLE=ISAMP
1340 POWER=SAMPLE/TOTL
1341 POWERI=1.0/POWER
1342 MN=0
1343 SUM =0.0
1345 CALL EXPVAR2(X,RANDOM,KZ)
1346 SUM =SUM+X
1347 CLASS(I)=(1.0+SUM *(ALPHA-1.0)/BETA)**(1.0/(1.0-ALPHA))
1348 TEST=EXPF(-CLASS(I))
1349 CALL UNIFORM(X,RANDOM,KZ)
1350 IF(TEST-X) 1345,1351,1351
1351 IF(CLASS(I)-POWERI) 1354,1352,1352
1352 I=I+1
1353 GOTO 1345
```

ESTIMATE OF K_S (lines 1336–1342)

FINDING INTENSITIES $\frac{1}{K_S} \le X_I < 1.0$ (lines 1343–1353)

```
1354 SUM3=CLASS(I)**(2.0-ALPHA)
1355 JZ=I
1356 SUM =0.0
1357 I=I+1
1359 CALL EXPVAR2(X,RANDOM,KZ)
1360 SUM =SUM+X
1361 TOTL=SUM3-SUM *(2.0-ALPHA)/(BETA*POWER)
1362 IF(TOTL) 1370,1363,1363
1363 CLASS(I)=TOTL**(1.0/(2.0-ALPHA))
1364 TEST=EXPF(-CLASS(I))
1365 CALL UNIFORM(X,RANDOM,KZ)
1366 IF(TEST-X) 1359,1357,1357
```

FINDING INTENSITIES $0 \leq X_i < \frac{1}{K_s}$

```
1370 DO 1376 LZ=1,JZ
1371 PMEAN=POWER*CLASS(LZ)
1372 IF(PMEAN-6.0) 1373,1375,1375
1373 CALL SMPOSN(PMEAN,INCLS(LZ),RANDOM,KZ)
1374 GOTO 1376
1375 CALL POISSON(PMEAN,INCLS(LZ),RANDOM,KZ,PSIII)
1376 CONTINUE
1377 LZ1=I-1
1378 LLZ=JZ+1
1379 DO 1387 LZ=LLZ,LZ1
1380 PMEAN=POWER*CLASS(LZ)
1381 TEST=(1.0-EXPF(-PMEAN))/PMEAN
1382 CALL UNIFORM(X,RANDOM,KZ)
```

FINDING NUMBER OF INDIVIDUALS IN EACH SPECIES

```
1383 IF(TEST-X) 1384,1386,1386
1384 INCLS(LZ)=0
1385 GOTO 1387
1386 CALL TRUNCP(PMEAN,INCLS(LZ),RANDOM,KZ)
1387 CONTINUE
     IF(MN) 1405,1388,1405
1388 INDIVS=0
1389 DO 1391 LZ=1,LZ1
1390 INDIVS=INDIVS+INCLS(LZ)
1391 INCL2(LZ)=INCLS(LZ)
     INDIVS1=INDIVS
     IDSAMP=ISAMP
1392 IF(INDIVS-ISAMP) 1393,1430,1415
1393 IDSAMP=ISAMP-INDIVS
     MN=1
1394 DSAMP=IDSAMP
1395 DINDV=INDIVS
1396  POWER=2.0*DSAMP*POWER/DINDV
1397 JQ=JZ
1398 JZ=I-1
1400 POWERI=1.0/(1.0/POWERI+POWER)
1401 IF(CLASS(JQ)-POWERI) 1356,1356,1410
1405 INDIVS1=0
1406 DO 1407 LZ=1,LZ1
1407 INDIVS1=INDIVS1+INCLS(LZ)
```

ADD EXTRA INDIVIDUALS WHEN NECESSARY

```
1408 IF(INDIVS1-IDSAMP) 1409,1430,1415
1409 INDIVS=INDIVS+INDIVS1
1410 POWER=1.0/POWERI
1411 GOTO 1393
1415 N3=INDIVS1-IDSAMP
1416 II=0
1417 II=II+1
1418 IF(INCLS(II)) 1417,1417,1421
1421 IF(INCLS(II)*N3-7*INDIVS1) 1422,1422,1424
1422 CALL HYPER(INDIVS1,N3,INCLS(II),NH,RANDOM,KZ,PSIII)
1423 GOTO 1425
1424 CALL HYPERL(INDIVS1,N3,INCLS(II),NH,RANDOM,KZ,PSIII)
1425 INDIVS1=INDIVS1-INCLS(II)
1426 N3=N3-NH
1427 INCLS(II)=INCLS(II)-NH
1428 IF(N3) 1430,1430,1417
1430 IF(MN) 1433,1433,1431
1431 DO 1432 LZ=1,LZ1
1432 INCLS(LZ)=INCL2(LZ)+INCLS(LZ)
1433 DO 1434 LZ=1,100
1434 INCL2(LZ)=0
```

ELIMINATE EXTRA INDIVIDUALS WHEN NECESSARY

```
1435 JA=100
1436 DO 1443 LZ=1,LZ1
1437 KKK=INCLS(LZ)
1438 IF(KKK-100) 1439,1439,1441
```

PRINT SAMPLE

```
1439 INCL2(KKK)=INCL2(KKK)+1
1440 GOTO 1443
1441 JA=JA+1
1442 INCL2(JA)=KKK
1443 CONTINUE
1444 WRITE OUTPUT TAPE3,1481,(INCL2(K),K=1,100)
     IF(JA-101) 1446,1445,1445
1445 WRITE OUTPUT TAPE3,1482,(INCL2(K),K=101,JA)
1446 STOP 0007
     END
     SUBROUTINE HYPER(NTOT,N3,NDEF,NH,RANDOM,KZ,PSIII)
     DIMENSION RANDOM(72),PSIII(11)
 401 PARAM1=N3
 402 PARAM2=NTOT-NDEF-N3+1
 403 PARAM=PARAM1/PARAM2
     PARAM=PARAM/(1.0+PARAM)
 404 CALL SMBIN(NDEF,PARAM,NH,RANDOM,KZ)
 405 IF(1-NH) 406,435,435
 406 X=NH
 407 Y=X-0.5
 409 U1=PARAM1-Y
 410 CALL PSII(U1,PSIFU,PSIII)
 412 U2=PARAM2+Y
 413 CALL PSII(U2,PSIFU1,PSIII)
 414 U1=-Y/PARAM1
```

FINDING HYPERGEOMETRIC RANDOM VARIABLE USING ACCEPTANCE REJECTION PROCEDURE WITH BINOMIAL

```
415 CALL SIMPL(U1,PROB1)
416 U2=Y/PARAM2
417 CALL SIMPL(U2,PROB2)
418 PROB1=-PARAM1*PROB1-PARAM2*PROB2                    +PSIFU1+PSIFU
425 CALL EXPVAR2(TEST,RANDOM,KZ)
426 RAND=TEST+PROB1
427 IF( RAND) 428,429,429
428 GOTO 401
429 CONTINUE
435 CONTINUE
    RETURN
    END
    SUBROUTINE SMPOSN(PARAM,NP,RANDOM,KZ)
    DIMENSION RANDOM(72)
450 NP=-1
451 SUM=0.0
452 NP=NP+1
453 CALL EXPVAR2(X,RANDOM,KZ)
454 SUM=SUM+X
455 IF(SUM-PARAM) 452,452,456
456 CONTINUE
    RETURN
    END
    SUBROUTINE TRUNCP(PARAM,NPT,RANDOM,KZ)
    DIMENSION RANDOM(72)
```

FINDING POISSON RANDOM VARIABLE
USING METHOD OF ADDING
EXPONENTIAL RANDOM VARIABLES

```
469 SUM=0.0
470 NPT=0
471 CALL EXPVAR2(X,RANDOM,KZ)
472 IQT=X/PARAM
473 QT=IQT+1
474 SUM=X
475 PARAM=QT*PARAM
476 NPT=NPT+1
477 CALL EXPVAR2(X,RANDOM,KZ)
478 SUM=SUM+X
479 IF(SUM-PARAM) 476,476,480
480 CONTINUE
    RETURN
    END
```

FINDING TRUNCATED POISSON RANDOM VARIABLE

```
    SUBROUTINE POISSON(PARAM,NP,RANDOM,KZ,PSIII)
    DIMENSION RANDOM(72),PSIII(11)
    MODE=PARAM
    J=1
    INTP=SQRTF(2.0*PARAM)
    PN1=MODE+INTP
    ALPHA=LOGF((PN1+1.0)/PARAM)
    PN2=MODE-INTP-1
    BETA=LOGF(PARAM/(PN2+1.0))
    NTOT=0
    N3=0
```

CALCULATE POISSON PARAMETERS FOR FIT SUBROUTINE

```
NPAR=0
NDEF=0
CALL FIT(MODE,INTP,ALPHA,BETA,NP,J,PARAM,NPAR,NDEF,N3,NTOT,
1RANDOM,KZ,PSIII)
RETURN
END
SUBROUTINE BINOML(NPAR,PARAM,NB,RANDOM,KZ,PSIII)
DIMENSION RANDOM(72),PSIII(11)
PARAMN=NPAR
MODE=(PARAMN+1.0)*PARAM
J=2
INTP=SQRTF(2.0*PARAMN*PARAM*(1.0-PARAM))
PNPAR=NPAR
PN1=MODE+INTP
ALPHA=LOGF((PN1+1.0)*(1.0-PARAM)/(PARAM*(PNPAR-PN1)))
PN2=MODE-INTP-1
BETA=LOGF(PARAM*(PNPAR-PN2)/((1.0-PARAM)*(PN2+1.0)))
NTOT=0
N3=0
NDEF=0
CALL FIT(MODE,INTP,ALPHA,BETA,NB,J,PARAM,NPAR,NDEF,N3,NTOT,
1RANDOM,KZ,PSIII)
RETURN
END
SUBROUTINE HYPERL(NTOT,N3,NDEF,NH,RANDOM,KZ,PSIII)
```

CALCULATE BINOMIAL PARAMETERS FOR FIT SUBROUTINE

```
DIMENSION RANDOM(72),PSIII(11)
PARAM1=(NDEF+1)*(N3+1)
PARAM2=(NTOT+2)
J=3
MODE=PARAM1/PARAM2
PARAM1=N3*NDEF*(NTOT-NDEF)*(NTOT-N3)
PARAM2=(NTOT-1)*NTOT**2
INTP=SQRTF(2.0*PARAM1/PARAM2)
PNDEF=NDEF
PNTOT=NTOT
PN3=N3
PN1=MODE+INTP
ALPHA=LOGF((PN1+1.0)*(PNTOT-PNDEF-PN3+PN1+1.0)/((PNDEF-PN1)
1*(PN3-PN1)))
PN2=MODE-INTP-1
BETA=LOGF((PNDEF-PN2)*(PN3-PN2)/((PN2+1.0)*(PNTOT-PNDEF-PN3+PN2
1+1.0)))
NPAR=0
PARAM=0.0
CALL FIT(MODE,INTP,ALPHA,BETA,NH,J,PARAM,NPAR,NDEF,N3,NTOT,
1RANDOM,KZ,PSIII)
RETURN
END
SUBROUTINE FIT(MODE,INTP,ALPHA,BETA,NP,J,PARAM,NPAR,NDEF,N3,NTOT,
1RANDOM,KZ,PSIII)
```

CALCULATE HYPERGEOMETRIC PARAMETERS FOR FIT SUBROUTINE

ACCEPTANCE-REJECTION
PROCEDURE DESCRIBED IN

```
      DIMENSION RANDOM(72),PSIII(11)
      CALL LOGPRB(MODE,ELOGPM,J,PARAM,NPAR,NDEF,N3,NTOT,PSIII)
      N1=MODE+INTP
      CALL LOGPRB(N1,ELOGP1,J,PARAM,NPAR,NDEF,N3,NTOT,PSIII)
      N2=MODE-INTP-1
      NZ=N2+1
      CALL LOGPRB(NZ,ELOGP2,J,PARAM,NPAR,NDEF,N3,NTOT,PSIII)
      PN1=N1
      KLOG1=(ELOGPM-ELOGP1)/ALPHA
      PN2=N2
      KLOG2=(ELOGPM-ELOGP2)/BETA
      FLTK1=KLOG1
      FLTK2=KLOG2
      U1=ELOGPM-ELOGP1-FLTK1*ALPHA
      U2=ELOGPM-ELOGP2-FLTK2*BETA
      FLTN=N1-N2-KLOG1-KLOG2-2
      TOT1=EXPF(-U1)/(1.0-EXPF(-ALPHA))
      TOT2=EXPF(-U2)/(1.0-EXPF(-BETA))
      TOT=TOT1+TOT2+FLTN
      TOT1=TOT1/TOT
      TOT2=TOT2/TOT
      FLTN=FLTN/TOT
  198 CALL UNIFORM(X,RANDOM,KZ)
      IF(X-TOT2-FLTN) 201,201,206
  201 IF(TOT2-X) 202,202,211
```

ACCEPTANCE-REJECTION PROCEDURE DESCRIBED IN CHAPTER 3 FOR FINDING RANDOM VARIABLES WITH SPECIFIED DISTRIBUTIONS

```
202 NS=(X-TOT2)*TOT
203 NS=N2+KLOG2+2+NS
204 PN=0.0
    ELOGP2=ELOGPM
205 GOTO 220
206 CALL EXPVAR(ALPHA,NS,RANDOM,KZ)
207 NS=N1-KLOG1+NS
208 PN=NS
209 PN=(PN-PN1)*ALPHA
    ELOGP2=ELOGP1
210 GOTO 220
211 CALL EXPVAR(BETA,NS,RANDOM,KZ)
212 NS=N2+KLOG2-NS+1
213 PN=NS
214 PN=(PN2-PN+1.0)*BETA
220 CALL LOGPRB(NS,ESTPN,J,PARAM,NPAR,NDEF,N3,NTOT,PSIII)
221 CALL EXPVAR2(TEST,RANDOM,KZ)
222 RAND=TEST-(ELOGP2-PN-ESTPN)
223 IF( RAND) 224,225,225
224 GOTO 198
225 NP=NS
    RETURN
    END
    SUBROUTINE LOGPRB (NS,PROB,J,PARAM,NPAR,NDEF,N3,NTOT,PSIII)
    DIMENSION PSIII(11)
```

```
      IF(J-2) 300,310,330
  300 X=NS
      IF(NS) 365,301,301
  301 Y=X+0.5                                              } CALCULATE LOGARITHM OF
  302 CALL PSII(Y,PSIFU,PSIII)                             } POISSON PROBABILITY
  303 U2=(X-PARAM+0.5)/PARAM
  304 CALL SIMPL(U2,PROB)
  305 PROB=-PARAM*(PROB-1.0)+PSIFU
      RETURN
  310 IF(NS-NPAR)311,311,365
  311 IF(-NS) 312,312,365
  312 X=NS
  313 Y=NPAR+1
  314 U2=(X+0.5-Y*PARAM)/(Y*PARAM)                         } CALCULATE LOGARITHM OF
  315 CALL SIMPL(U2,PROB)                                  } BINOMIAL PROBABILITY
  316 U3=-(X+0.5-Y*PARAM)/(Y*(1.0-PARAM))
  317 CALL SIMPL(U3,PROB1)
  319 X=X+0.5
  320 CALL PSII(X,PSIFU,PSIII)
  321 X=Y-X
  322 CALL PSII(X,PSIFU1,PSIII)
  323 PROB=-Y*PARAM*PROB-Y*(1.0-PARAM)*PROB1+PSIFU+PSIFU1        -0.5*
     1LOGF(Y*PARAM*(1.0-PARAM))
      RETURN
  330 IF(NS-NDEF) 331,331,365
```

```
331 IF(NS-N3) 332,332,365
332 IF(-NS) 333,333,365
333 X=NS
334 PARAM1=NDEF+1
335 PARAM2=N3+1
336 Y=NTOT+2
337 X=X+0.5
338 U1=PARAM2*PARAM1/Y
339 Y1=X-U1
340 U2=Y1/U1
341 CALL SIMPL(U2,PROB1)
342 U2=-Y1/(PARAM2-U1)
343 CALL SIMPL(U2,PROB2)
344 U2=-Y1/(PARAM1-U1)
345 CALL SIMPL(U2,PROB3)
346 U2=Y1/(Y-PARAM1-PARAM2+U1)
347 CALL SIMPL(U2,PROB4)
348 CALL PSII(X,PSIFU1,PSIII)
349 PARAM2=PARAM2-X
350 CALL PSII(PARAM2,PSIFU2,PSIII)
351 PARAM1=PARAM1-X
352 CALL PSII(PARAM1,PSIFU3,PSIII)
353 Y=Y-PARAM1-PARAM2-X
354 CALL PSII(Y,PSIFU4,PSIII)
355 PARAM1=PARAM1+Y1
```

CALCULATE LOGARITHM OF HYPERGEOMETRIC PROBABILITY

```
356 PARAM2=PARAM2+Y1
357 Y=Y-Y1
360 PROB=U1*PROB1+PARAM2*PROB2+PARAM1*PROB3+Y*PROB4-PSIFU1-PSIFU2
   1-PSIFU3-PSIFU4
361 PROB=-PROB
    RETURN
365 PROB=-10.0**10
    RETURN
    END
    SUBROUTINE SIMPL(X,PROB)
380 Y=X/(2.0+X)
381 IF(ABSF(Y)- 0.1        ) 382,382,385
382 Y2=Y**2
383 PROB=X*Y+(1.0+X)*2.*(((7.0*Y2+9.0)*Y2+12.6)*Y2+21.0)*Y2*Y/63.0
384 RETURN
385 PROB=X*Y+(1.0+X)*(LOGF(1.0+X)-2.0*Y)
    RETURN
    END
    SUBROUTINE EXPVAR2(X,RANDOM,KZ)
    DIMENSION RANDOM(72)
    CALL UNIFORM(X,RANDOM,KZ)
    X=-LOGF(X)
    RETURN
    END
    SUBROUTINE PSII(X,PROB,PSIII)
```

CALCULATE $(1+x)\log(1+x)-x$ (lines 380–END)

FINDING EXPONENTIAL RANDOM VARIABLE (SUBROUTINE EXPVAR2)

```
      DIMENSION PSIII(11)
  501 IF(X-11.0) 502,502,504
  502 KS=X+1.0
  503 PROB=PSIII(KS)
      RETURN
  504 X2=1.0/X**2
  505 PROB=((((+0.0008401068*X2-0.00059058779)*X2+0.00076686492)*X2
     1-0.0024305555)*X2+0.0416666666)/X
      RETURN
      END
```

CALCULATE $\psi(x)$

```
      SUBROUTINE SMBIN(NPAR,PARAM,NB,RANDOM,KZ)
      DIMENSION RANDOM(72)
  550 PAR=NPAR
  551 PARAM=-LOGF(1.0-PARAM)
  552 QQ=PAR*PARAM
  553 NB=0
  554 S=-QQ
  556 CALL EXPVAR2(X,RANDOM,KZ)
  557 S=S+X
  558 IF(S) 559,567,567
  559 IF(X-PARAM) 560,560,565
  560 IF(NB) 561,565,561
  561 REM=(QQ+S)/PARAM
  562 IREM=REM
      REM1=IREM
```

FINDING BINOMIAL RANDOM VARIABLE BY THE METHOD OF ADDING EXPONENTIAL RANDOM VARIABLES

```
563 REM=REM-REM1
564 IF(REM-X/PARAM) 565,556,556
565 NB=NB+1
566 GOTO 556
567 CONTINUE
    RETURN
    END
```

```
    SUBROUTINE EXPVAR(PARAM,NS,RANDOM,KZ)
    DIMENSION RANDOM(72)
    CALL UNIFORM(X,RANDOM,KZ)
    NS=-LOGF(X)/PARAM
    RETURN
    END
```

FINDING DISCRETE EXPONENTIAL RANDOM VARIABLE

```
    SUBROUTINE RANREAD(RANDOM,KZ)
    DIMENSION MANY(54),RANDOM(72)
    CON(KK=00007777777777777B)
    READ TAPE 1, (MANY(L),L=1,54)
    KZ=1
    LCYC=18
    I=0
    ENI1(0) ENI2(C)
 1K ENA(0) LDQ1(MANY+1)
    LLS(12) LDQ1(MANY+2)
    LLS(12) LDQ1(MANY+3)
    LLS(12) ENI4(3)
```

FORTRAN SYMBOLIC PROGRAM TO OBTAIN UNIFORM RANDOM VARIABLES FROM BINARY RANDOM BITS.

```
      SLJ0(3K)
 2K   INI1(1) LDG1(MANY)
      LDL(KK)
 3K   SCA3(2012B) LRS(47)
      ENA3(76000B) AJP1(4K)
      INI3(1) ENI(0)
 4K   ENA3(0) LLS(36)
      INI2(1) STA2(RANDOM)
      IJP4(2K) ENI(0)
 5K   RSO(LCYC) AJP1(1K)
      RETURN
      END
      SUBROUTINE UNIFORM(X,RANDOM,KZ)
      DIMENSION RANDOM(72)
589   FORMAT(I5)
590   KZ=KZ+1
591   IF(KZ-72) 593,593,592
592   CALL RANREAD(RANDOM,KZ)
      WRITE OUTPUT TAPE3,589,KZ
593   X=RANDOM(KZ)
      IF(X-10.0**(-15.0)) 590,590,594
594   CONTINUE
      RETURN
      END
      END
```

FINDING UNIFORM RANDOM VARIABLE

```
      PROGRAM SPECIES 4
      DIMENSION INCLS(2000),PSIII(11),RANDOM(72)
 1480 FORMAT(I5,2X,E13.5)
 1481 FORMAT(10(I5))
 1482 FORMAT(2X,I5)
      PAUSE 0001
 1200 X=1.0/10.5**2
 1201 PSIII(11)=((((+0.0008401068*X-0.00059058779)*X+0.0007688492)*X
     1-0.00243055555)*X+0.0416666666)/10.5
 1202 DO 1206 J=1,10
 1203 PNI=11-J
 1204 PNI=1.0/(2.0*PNI)**2
 1205 DELTA=((((((PNI/210.0+1.0/156.0)*PNI+1.0/110.0)*PNI+1.0/72.0)*
     1PNI+1.0/42.0)*PNI+0.03)*PNI+1.0/6.0)*PNI
 1206 PSIII(11-J)=PSIII(12-J)-DELTA
 1207 PSIII(1)=PSIII(1)-(((0.25/506.0+1.0/421.0)*0.25+1.0/342.0)*0.25
     1+1.0/272.0)*0.25**8
 1210 CALL RANREAD(RANDOM,KZ)
 1499 DO 1500 J=1,100
 1500 INCLS(J)=0
 1501 READ INPUT TAPE2,1480,ISAMP,PARAM
 1502 I=101
 1504 CALL UNIFORM(X,RANDOM,KZ)
 1505 Y=1.0-X**(1.0/PARAM)
      ISAMP=ISAMP-1
```

} **CALCULATE ψ (i) i = 1,2,...11** (statements 1200–1207)

ISAMP = N

PARAM = A

} **FIND RANDOM VARIABLE WITH DENSITY $f(x) = A(1-x)^{A-1}$ $0 \leq x \leq 1$** (statements 1504–1505)

```
      IF(ISAMP) 1524,1524,1506
1506 SAMPLE=ISAMP
1507 AG=SAMPLE*Y
1508 IF(AG-8.0) 1510,1512,1512                          OBTAIN BINOMIAL RANDOM
1510 CALL SMBIN(ISAMP,Y,KKK,RANDOM,KZ)                  VARIABLE
1511 GOTO 1513                                           KKK = NUMBER OF INDIVIDUALS IN
1512 CALL BINOML(ISAMP,Y,KKK,RANDOM,KZ,PSIII)            THIS SPECIES
1513 KKK=KKK+1
1514 IF(KKK-100) 1515,1515,1516
1515 INCLS(KKK)=INCLS(KKK)+1
1516 ISAMP=ISAMP-KKK+1
1517 IF(ISAMP) 1525,1525,1504
1518 INCLS(I)=KKK
1519 I=I+1
1520 IF(I-2000) 1516,1521,1521                           STOP IF 2000 SPECIES HAVE BEEN
1521 STOP 0003                                           FOUND
1524 INCLS(1)=INCLS(1)+1
1525 WRITE OUTPUT TAPE3,1481,(INCLS(K),K=1,100)        }
      IF(I-101) 1527,1527,1522                         }
1522 I=I-1                                             }
1526 WRITE OUTPUT TAPE3,1482,(INCLS(K),K=101,I)        } PRINT SAMPLE
1527 STOP 0004
      END
      SUBROUTINE BINOML(NPAR,PARAM,NB,RANDOM,KZ,PSIII)
      DIMENSION RANDOM(72),PSIII(11)
```

ALL FOLLOWING SUBROUTINES THE
SAME AS THOSE IN PROGRAM

ALL FOLLOWING SUBROUTINES THE
SAME AS THOSE IN PROGRAM
SPECIES 3

```
      PARAMN=NPAR
      MODE=(PARAMN+1.0)*PARAM
      J=2
      INTP=SQRTF(2.0*PARAMN*PARAM*(1.0-PARAM))
      PNPAR=NPAR
      PN1=MODE+INTP
      ALPHA=LOGF((PN1+1.0)*(1.0-PARAM)/(PARAM*(PNPAR-PN1)))
      PN2=MODE-INTP-1
      BETA=LOGF(PARAM*(PNPAR-PN2)/((1.0-PARAM)*(PN2+1.0)))
      NTOT=0
      N3=0
      NDEF=0
      CALL FIT(MODE,INTP,ALPHA,BETA,NB,J,PARAM,NPAR,NDEF,N3,NTOT,
     1RANDOM,KZ,PSIII)
      RETURN
      END
      SUBROUTINE FIT(MODE,INTP,ALPHA,BETA,NP,J,PARAM,NPAR,NDEF,N3,NTOT,
     1RANDOM,KZ,PSIII)
      DIMENSION RANDOM(72),PSIII(11)
      CALL LOGPRB(MODE,ELOGPM,J,PARAM,NPAR,NDEF,N3,NTOT,PSIII)
      N1=MODE+INTP
      CALL LOGPRB(N1,ELOGP1,J,PARAM,NPAR,NDEF,N3,NTOT,PSIII)
      N2=MODE-INTP-1
      NZ=N2+1
      CALL LOGPRB(NZ,ELOGP2,J,PARAM,NPAR,NDEF,N3,NTOT,PSIII)
```

```
      PN1=N1
      KLOG1=(ELOGPM-ELOGP1)/ALPHA
      PN2=N2
      KLOG2=(ELOGPM-ELOGP2)/BETA
      FLTK1=KLOG1
      FLTK2=KLOG2
      U1=ELOGPM-ELOGP1-FLTK1*ALPHA
      U2=ELOGPM-ELOGP2-FLTK2*BETA
      FLTN=N1-N2-KLOG1-KLOG2-2
      TOT1=EXPF(-U1)/(1.0-EXPF(-ALPHA))
      TOT2=EXPF(-U2)/(1.0-EXPF(-BETA))
      TOT=TOT1+TOT2+FLTN
      TOT1=TOT1/TOT
      TOT2=TOT2/TOT
      FLTN=FLTN/TOT
  198 CALL UNIFORM(X,RANDOM,KZ)
      IF(X-TOT2-FLTN) 201,201,206
  201 IF(TOT2-X) 202,202,211
  202 NS=(X-TOT2)*TOT
  203 NS=N2+KLOG2+2+NS
  204 PN=0.0
      ELOGP2=ELOGPM
  205 GOTO 220
  206 CALL EXPVAR(ALPHA,NS,RANDOM,KZ)
  207 NS=N1-KLOG1+NS
```

```
208 PN=NS
209 PN=(PN-PN1)*ALPHA
    ELOGP2=ELOGP1
210 GOTO 220
211 CALL EXPVAR(BETA,NS,RANDOM,K2)
212 NS=N2+KLOG2-NS+1
213 PN=NS
214 PN=(PN2-PN+1.0)*BETA
220 CALL LOGPRB(NS,ESTPN,J,PARAM,NPAR,NDEF,N3,NTOT,PSIII)
221 CALL EXPVAR2(TEST,RANDOM,K2)
222 RAND=TEST-(ELOGP2-PN-ESTPN)
223 IF( RAND) 224,225,225
224 GOTO 198
225 NP=NS
    RETURN
    END
    SUBROUTINE LOGPRB (NS,PROB,J,PARAM,NPAR,NDEF,N3,NTOT,PSIII)
    DIMENSION PSIII(11)
    IF(J-2) 300,310,330
300 X=NS
    IF(NS) 365,301,301
301 Y=X+0.5
302 CALL PSII(Y,PSIFU,PSIII)
303 U2=(X-PARAM+0.5)/PARAM
304 CALL SIMPL(U2,PROB)
```

```
305 PROB=-PARAM*(PROB-1.0)+PSIFU
    RETURN
310 IF(NS-NPAR)311,311,365
311 IF(-NS) 312,312,365
312 X=NS
313 Y=NPAR+1
314 U2=(X+0.5-Y*PARAM)/(Y*PARAM)
315 CALL SIMPL(U2,PROB)
316 U3=-(X+0.5-Y*PARAM)/(Y*(1.0-PARAM))
317 CALL SIMPL(U3,PROB1)
319 X=X+0.5
320 CALL PSII(X,PSIFU,PSIII)
321 X=Y-X
322 CALL PSII(X,PSIFU1,PSIII)
323 PROB=-Y*PARAM*PROB-Y*(1.0-PARAM)*PROB1+PSIFU+PSIFU1        -0.5*
   1LOGF(Y*PARAM*(1.0-PARAM))
    RETURN
330 IF(NS-NDEF) 331,331,365
331 IF(NS-N3) 332,332,365
332 IF(-NS) 333,333,365
333 X=NS
334 PARAM1=NDEF+1
335 PARAM2=N3+1
336 Y=NTOT+2
337 X=X+0.5
```

```
338 U1=PARAM2*PARAM1/Y
339 Y1=X-U1
340 U2=Y1/U1
341 CALL SIMPL(U2,PROB1)
342 U2=-Y1/(PARAM2-U1)
343 CALL SIMPL(U2,PROB2)
344 U2=-Y1/(PARAM1-U1)
345 CALL SIMPL(U2,PROB3)
346 U2=Y1/(Y-PARAM1-PARAM2+U1)
347 CALL SIMPL(U2,PROB4)
348 CALL PSII(X,PSIFU1,PSIII)
349 PARAM2=PARAM2-X
350 CALL PSII(PARAM2,PSIFU2,PSIII)
351 PARAM1=PARAM1-X
352 CALL PSII(PARAM1,PSIFU3,PSIII)
353 Y=Y-PARAM1-PARAM2-X
354 CALL PSII(Y,PSIFU4,PSIII)
355 PARAM1=PARAM1+Y1
356 PARAM2=PARAM2+Y1
357 Y=Y-Y1
360 PROB=U1*PROB1+PARAM2*PROB2+PARAM1*PROB3+Y*PROB4-PSIFU1-PSIFU2
   1-PSIFU3-PSIFU4
361 PROB=-PROB
    RETURN
365 PROB=-10.0**10
```

```
      RETURN
      END
      SUBROUTINE SIMPL(X,PROB)
  380 Y=X/(2.0+X)
  381 IF(ABSF(Y)- 0.1        ) 382,382,385
  382 Y2=Y**2
  383 PROB=X*Y+(1.0+X)*2.*(((7.0*Y2+9.0)*Y2+12.6)*Y2+21.0)*Y2*Y/63.0
  384 RETURN
  385 PROB=X*Y+(1.0+X)*(LOGF(1.0+X)-2.0*Y)
      RETURN
      END
      SUBROUTINE EXPVAR2(X,RANDOM,KZ)
      DIMENSION RANDOM(72)
      CALL UNIFORM(X,RANDOM,KZ)
      X=-LOGF(X)
      RETURN
      END
      SUBROUTINE PSII(X,PROB,PSIII)
      DIMENSION PSIII(11)
  501 IF(X-11.0) 502,502,504
  502 KS=X+1.0
  503 PROB=PSIII(KS)
      RETURN
  504 X2=1.0/X**2
  505 PROB=((((+0.000840106 8*X2-0.00059058779)*X2+0.00076833492)*X2
```

```
    1-0.00243055555)*X2+0.0416666666)/X
      RETURN
      END
      SUBROUTINE SMBIN(NPAR,PARAM,NB,RANDOM,KZ)
      DIMENSION RANDOM(72)
  550 PAR=NPAR
  551 PARAM=-LOGF(1.0-PARAM)
  552 QQ=PAR*PARAM
  553 NB=0
  554 S=-QQ
  556 CALL EXPVAR2(X,RANDOM,KZ)
  557 S=S+X
  558 IF(S) 559,567,567
  559 IF(X-PARAM) 560,560,565
  560 IF(NB) 561,563,561
  561 REM=(QQ+S)/PARAM
  562 IREM=REM
      REM1=IREM
  563 REM=REM-REM1
  564 IF(REM-X/PARAM) 565,556,556
  565 NB=NB+1
  566 GOTO 556
  567 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE EXPVAR(PARAM,NS,RANDOM,KZ)
      DIMENSION RANDOM(72)
      CALL UNIFORM(X,RANDOM,KZ)
      NS=-LOGF(X)/PARAM
      RETURN
      END
      SUBROUTINE RANREAD(RANDOM,KZ)
      DIMENSION MANY(54),RANDOM(72)
      CON(KK=000C7777777777775)
      READ TAPE 1, (MANY(L),L=1,54)
      KZ=1
      LCYC=18
      I=0
      ENI1(0) ENI2(0)
   1K ENA(0) LDG1(MANY+1)
      LLS(12) LDG1(MANY+2)
      LLS(12) LDG1(MANY+3)
      LLS(12) ENI4(3)
      SLJ0(3K)
   2K INI1(1) LDG1(MANY)
      LDL(KK)
   3K SCA3(20125) LRS(47)
      ENA3(76000B) AJP1(4K)
      INI3(1) ENI(0)
   4K ENA3(0) LLS(36)
```

```
      INI2(1) STA2(RANDOM)
      IJP4(2K) ENI(0)
   5K RSO(LCYC) AJP1(1K)
      RETURN
      END
      SUBROUTINE UNIFORM(X,RANDOM,KZ)
      DIMENSION RANDOM(72)
  589 FORMAT(I5)
  590 KZ=KZ+1
  591 IF(KZ-72) 593,593,592
  592 CALL RANREAD(RANDOM,KZ)
      WRITE OUTPUT TAPE3,589,KZ
  593 X=RANDOM(KZ)
      RETURN
      END
      END
```

BIBLIOGRAPHY

[1] BIERENS DEHAAN, DAVID
Nouvelles Tables D'Intégrales Défines G. E. Stechert & Co. New York 1939.

[2] CARATHÉODORY, C.
Theory of Functions, Volume I, Part Five, Chelsea, New York 1958.

[3] DAVIS, HAROLD T.
Tables of the Higher Mathematical Functions, Volume I, Principia Press Bloomington, Indiana 1933.

[4] FISHER, R. A., A.S. CORBET and C. B. WILLIAMS
The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. Journal of Animal Ecology, Volume 12(1943) pp.42-58.

[5] RUBIN, HERMAN
Construction of Random Variables with Specified Distributions, Research Memorandum #88, Department of Statistics, Michigan State University, 1961.

[6] WILLIAMS, C. B.
Some Application of the Logarithmic Series and the Index of Diversity to Ecological Problems. Journal of Ecology, Volume 32(1944) pp.1-44.