Ann. Inst. Statist. Math.
Vol. 45, No. 2, 201–210 (1993)

ON A GENERALIZATION OF MORISITA'S MODEL FOR ESTIMATING THE HABITAT PREFERENCE

CH. A. CHARALAMBIDES AND M. V. KOUTRAS

Department of Mathematics, Section of Statistics and O.R., University of Athens, Panepistemiopolis, Athens 157 84, Greece

(Received January 27, 1992; revised July 6, 1992)

Abstract. The Morisita's model for estimating the habitat preference by the ant lions *Genuroides japonicus* is generalized by introducing, in addition to the environmental densities a and b, a repulsivity parameter θ. The probability function of the number L_n of individuals choosing fine sand to settle when a total of n ant lions are introduced is examined. A heuristic and the minimum chi-square methods for estimating the parameters a, b and θ are discussed.

Key words and phrases: Eulerian numbers, Gould and Hopper numbers, minimum chi-square estimation, unimodality.

1. Introduction

Morisita (1971), after a series of experimental studies on the habitat preference by the ant lions *Genuroides japonicus*, introduced the concept of "environmental density", which gives the value of a habitat expressing its unfavorableness for settling of an animal which has a strong mutual-repulsive influence to other individuals in the environment. According to Morisita, the ant lions have a strong tendency to prefer fine to coarse sand for pit formation when the population density is low; this tendency gradually falls with increasing density until an almost equal number of individuals settles in both sands.

Morisita's experimental data were collected by placing a total of n ant lions ($n = 1, 2, \dots, 7$) at the center of a box containing fine sand in one half and coarse sand in the other half. When all n ant lions had settled (by digging a small pit somewhere in the box), the number of individuals in the fine sand and coarse sand were counted. Morisita modelled his experiments assuming that

$$\begin{aligned}
&\Pr(\text{the first ant lion to choose coarse sand})\\
&\qquad = 1 - \Pr(\text{the first ant lion to choose fine sand}) = a/(a+b),\\
&\Pr(\text{the } (n+1)\text{-st ant lion to choose coarse sand given that}\\
&\qquad k \text{ ant lions are in fine sand}) = (a+k)/(a+b+n),
\end{aligned}$$

Pr(the $(n+1)$-st ant lion to choose fine sand given that
$n-k$ ant lions are in coarse sand) $= (b+n-k)/(a+b+n)$,

where the positive real parameters a and b expressing the "degree of unfavorableness" of fine and coarse sand respectively, are the aforementioned *environmental densities*. Consider the number L_n of ant lions choosing fine sand to settle and let $p_k(n;a,b) = \Pr(L_n = k)$, $k = 0,1,\ldots,n$. Morisita (1971) deduced a recurrence relation for $p_k(n;a,b)$, $k = 0,1,\ldots,n$ which was subsequently used for the estimation of the unknown parameters a and b. Janardan (1988) gave an explicit solution for $p_k(n,a,b)$ in terms of the generalized Eulerian numbers $E_{n,k}(a,b)$ and estimated a and b utilizing the methods of moments and maximum likelihood. It should also be noted that, as Charalambides (1991) indicated, $E_{n,k}(a,b)$ are directly related to the (symmetric) generalized Eulerian numbers of Carlitz and Scoville (1974).

In the present paper we consider (Section 2) a certain generalization of Morisita's model by introducing an additional parameter θ whose purpose is to give the model more flexibility in the explanation of individuals' repulsive behaviour. For the study of the resulting probability function $p_k(n;a,b,\theta)$ we employ proper generalizations of the Eulerian numbers of the binomial coefficients which are introduced and extensively studied in Section 3. In Section 4 we give formulas for the probability generating function and factorial moments of the distribution $p_k(n;a,b,\theta)$. Finally, in Section 5, we indicate how our model can be used for improving the fitting of Morisita's experimental data.

2. A generalized Morisita model

In Morisita's model the $(n+1)$-st ant lion selects fine (coarse) sand with probability proportional to $a+nx$ $(b+ny)$ where a, b are the environmental densities and x, y are the repulsivity factors, expressing the unfavorableness due to the population present in habitat. Morisita assumed that the unfavorableness due to the presence of an individual is one unit and therefore $x = k/n$ $(y = 1-k/n)$ is the proportion of individuals settled in fine (coarse) sand respectively.

In our generalized model we introduce a new parameter θ, $(-a/n \le \theta \le b/n)$ called hereafter *repulsivity parameter*, and assume that the repulsivity factors x and y are given by

$$x = \frac{k}{n} + \theta, \qquad y = 1 - \frac{k}{n} - \theta.$$

So, our generalized model is completely described by the following postulates

(2.1) Pr(the first ant lion to choose coarse sand)
$= 1 -$ Pr(the first ant lion to choose find sand) $= a/(a+b)$,

(2.2) Pr(the $(n+1)$-st ant lion to choose coarse sand given that k ant lions are in fine sand) $= (a+k+n\theta)/(a+b+n)$,

(2.3) Pr(the $(n+1)$-st ant lion to choose fine sand given that $n-k$ ant lions are in coarse sand) $= (b+n-k-n\theta)/(a+b+n)$.

Let L_n denote the number of individuals choosing fine sand to settle, when a total of n ant lions are introduced in the experiment. Making use of assumptions (2.1), (2.2) and (2.3), we may easily deduce that the probability function

$$p_k(n;a,b,\theta) = \Pr(L_n = k), \qquad k = 0,1,\ldots,n$$

satisfies the recurrence relation

$$\begin{aligned}(2.4)\qquad p_k(n+1;a,b,\theta) &= \frac{k+a+\theta n}{a+b+n} p_k(n;a,b,\theta) \\ &\quad + \frac{b+n-k+1-\theta n}{a+b+n} p_{k-1}(n;a,b,\theta), \\ &\qquad k = 1,2,\ldots,n+1, \quad n = 0,1,\ldots\end{aligned}$$

with initial conditions

$$\begin{aligned}(2.5)\qquad p_0(n;a,b,\theta) &= \Pr[L_1 = 0] \prod_{j=1}^{n-1} \Pr[L_{j+1} = 0 \mid L_j = 0] \\ &= \frac{\binom{-a/\theta}{n}(-\theta)^n}{\binom{a+b+n-1}{n}}, \qquad n \geq 0.\end{aligned}$$

This probability function can be written in the form

$$(2.6)\qquad p_k(n;a,b,\theta) = \frac{B_{n,k}(a,b,s)}{\binom{a+b+n-1}{n} s^n}, \qquad k = 0,1,\ldots,n, \quad s = -1/\theta$$

where, by (2.4) and (2.5), the double sequence of numbers $B_{n,k}(a,b,s)$, $k = 0,1,\ldots,n$, $n = 0,1,2,\ldots$ satisfies the recurrence relation

$$\begin{aligned}(2.7)\qquad (n+1)B_{n+1,k}(a,b,s) &= [s(k+a)-n]B_{n,k}(a,b,s) \\ &\quad + [s(n-k+b+1)+n]B_{n,k-1}(a,b,s) \\ &\qquad k = 1,2,\ldots,n+1, \quad n = 0,1,2,\ldots\end{aligned}$$

with initial conditions

$$(2.8)\qquad B_{0,0}(a,b,s) = 1, \qquad B_{0,k}(a,b,s) = 0, \quad k = 1,2,\ldots.$$

These numbers, which are asymptotically connected with the generalized Eulerian numbers (see Janardan (1988), Charalambides (1991)), are examined in the next section.

3. The number $B_{n,k}(a,b,s)$

Consider, first, the polynomial

$$B_n(t;a,b,s) = \sum_{k=0}^{n} B_{n,k}(a,b,s)t^k \qquad n = 0,1,2,\ldots. \tag{3.1}$$

Then, by virtue of (2.7) and (2.8) the difference-differential equation

$$\begin{aligned}(n+1)B_{n+1}(t;a,b,s) = st(1-t)\frac{d}{dt}B_n(t;a,b,s)& \\ + [(sn+sb+n)t + (sa-n)]B_n(t;a,b,s)&, \\ n = 0,1,2,\ldots&\end{aligned}$$

with $B_0(t;a,b,s) = 1$, is readily obtained. Introducing the auxiliary functions

$$C_n(t;a,b,s) = t^{a-n/s}(1-t)^{-(n+a+b)}B_n(t;a,b,s), \qquad n = 0,1,2,\ldots, \tag{3.2}$$

the more manageable difference-differential equation

$$(n+1)C_{n+1}(t;a,b,s) = st^{1-1/s}\frac{d}{dt}C_n(t;a,b,s), \qquad n = 0,1,2,\ldots \tag{3.3}$$

with $C_0(t;a,b,s) = t^a(1-t)^{-(a+b)}$, is deduced. Multiplying it by u^n and summing for $n = 0,1,2,\ldots$ it follows that the generating function

$$C(t,u;a,b,s) = \sum_{n=0}^{\infty} C_n(t;a,b,s)u^n \tag{3.4}$$

satisfies the partial differential equation

$$\frac{\partial C(t,u;a,b,s)}{\partial u} - st^{1-1/s}\frac{\partial C(t,u;a,b,s)}{\partial t} = 0$$

with $C(t,0;a,b,s) = C_0(t;a,b,s) = t^a(1-t)^{-(a+b)}$. The general solution of this equation may easily be obtained in the form

$$C(t,u;a,b,s) = \Psi(u+t^{1/s})$$

where Ψ is a function to be determined. Putting $u = 0$ and since $C(t,0;a,b,s) = t^a(1-t)^{-(a+b)}$ if follows that

$$\Psi(t^{1/s}) = t^a(1-t)^{-(a+b)}.$$

Hence $\Psi(w) = w^{sa}(1-w^s)^{-(a+b)}$ and

$$C(t,u;a,b,s) = (u+t^{1/s})^{sa}[1-(u+t^{1/s})^s]^{-(a+b)}. \tag{3.5}$$

The generating function

$$B(t,u;a,b,s) = \sum_{n=0}^{\infty} B_n(t;a,b,s)u^n = \sum_{n=0}^{\infty}\sum_{k=0}^{n} B_{n,k}(a,b,s)t^k u^n \tag{3.6}$$

by virtue of (3.2) may be obtained in terms of the generating function (3.4) as

$$B(t,u;a,b,s) = t^{-a}(1-t)^{a+b}C(t,t^{1/s}(1-t)u;a,b,s).$$

Thus, by (3.5)

$$B(t,u;a,b,s) = [1+(1-t)u]^{sa}\left\{\frac{1-t}{1-t[1+(1-t)u]^s}\right\}^{a+b}. \tag{3.7}$$

Expanding this generating function into powers of u,

$$\begin{aligned} B(t,u;a,b,s) &= (1-t)^{a+b}\sum_{j=0}^{\infty}\binom{a+b+j-1}{j}t^j[1+(1-t)u]^{s(j+a)} \\ &= (1-t)^{a+b}\sum_{j=0}^{\infty}\binom{a+b+j-1}{j}t^j\sum_{n=0}^{\infty}\binom{s(j+a)}{n}(1-t)^n u^n \\ &= \sum_{n=0}^{\infty}\left\{(1-t)^{n+a+b}\sum_{j=0}^{\infty}\binom{a+b+j-1}{j}\binom{s(j+a)}{n}t^j\right\}u^n, \end{aligned}$$

it follows, by (3.6), that

$$B_n(t;a,b,s) = (1-t)^{n+a+b}\sum_{j=0}^{\infty}\binom{a+b+j-1}{j}\binom{s(j+a)}{n}t^j \tag{3.8}$$

and

$$\begin{aligned} B_{n,k}(a,b,s) &= \sum_{j=0}^{k}(-1)^{k-j}\binom{a+b+n}{k-j}\binom{a+b+j-1}{j}\binom{s(j+a)}{n} \\ &= \sum_{j=0}^{k}(-1)^{j}\binom{a+b+n}{j}\binom{a+b+k-j-1}{k-j}\binom{s(k-j+a)}{n}. \end{aligned} \tag{3.9}$$

The following properties of the numbers $B_{n,k}(a,b,s)$ are worth noting.

(a) Taking the limit, as $t \to 1$, of the generating function (3.7) it follows that

$$\lim_{t\to 1} B(t,u;a,b,s) = (1-su)^{-(a+b)} = \sum_{n=0}^{\infty}\binom{a+b+n-1}{n}s^n u^n,$$

implying

$$B_n(1;a,b,s) = \sum_{k=0}^{\infty} B_{n,k}(a,b,s) = \binom{a+b+n-1}{n} s^n. \tag{3.10}$$

(b) From (3.9) and since

$$\lim_{s\to\pm\infty} s^{-n}\binom{s(k-j+a)}{n} = \frac{(k-j+a)^n}{n!}$$

it follows that

$$\lim_{s\to\pm\infty} s^{-n} B_{n,k}(a,b,s) = A_{n,k}(b,n)/n!, \tag{3.11}$$

where

$$A_{n,k}(b,a) = \sum_{j=0}^{\infty}(-1)^j \binom{a+b+n}{j}\binom{a+b+k-j-1}{k-j}(k-j+a)^n$$

is the generalized Eulerian number (Janardan (1988), Charalambides (1991)) which is related to the generalized symmetric Eulerian number $A(r,s \mid a,b)$ of Carlitz and Scoville (1974) by

$$A_{n,k}(b,a) = A(k,n-k \mid b,a) = A(n-k,k \mid a,b).$$

(c) The generating function (3.7) may be rewritten in the form

$$B(t,u;a,b,s) = [1-(t-1)u]^{-sb}\left\{1 - \frac{[1-(t-1)u]^{-s}-1}{t-1}\right\}^{-a-b}$$

which, on using the Gould-Hopper numbers (Gould and Hopper (1962), Charalambides and Koutras (1983))

$$G(n,r;\alpha,\beta) = \frac{1}{r!}\sum_{k=0}^{\infty}(-1)^{r-k}\binom{r}{k}(\alpha k+\beta)_n \tag{3.12}$$

with generating function

$$\sum_{n=r}^{\infty} G(n,r;\alpha,\beta)u^n/n! = (1+u)^{\beta}[(1+u)^{\alpha}-1]^r/r!,$$

can be expressed as

$$\begin{aligned}
&B(t,u;a,b,s)\\
&\quad= \sum_{r=0}^{\infty}\binom{a+b+r-1}{r}[1-(t-1)u]^{-sb}\{[1-(t-1)u]^{-s}-1\}^r(t-1)^{-r}\\
&\quad= \sum_{r=0}^{\infty}\binom{a+b+r-1}{r}\sum_{n=r}^{\infty}(-1)^n r! G(n,r;-s,-sb)(t-1)^{n-r}u^n/n!\\
&\quad= \sum_{n=0}^{\infty}\left\{\frac{(-1)^n}{n!}\sum_{r=0}^{\infty}\binom{a+b+r-1}{r} r! G(n,r;-s,-sb)(t-1)^{n-r}\right\}u^n,
\end{aligned}$$

yielding

$$(3.13)\quad B_n(t;a,b,s) = \frac{(-1)^n}{n!}\sum_{r=0}^{n}\binom{a+b+r-1}{r} r!G(n,r;-s,-sb)(t-1)^{n-r}.$$

Finally note that

$$B_{n,k}(0,1,-s) \equiv B_{n,k}(-s) = \sum_{j=0}^{k}(-1)^j\binom{n+1}{j}\binom{s(k-j)+n-1}{n}$$

is the number of sequences $\{i_1, i_2, \ldots, i_n\}$ with $i_r \in \{1, 2, \ldots, s\}$, $r = 1, 2, \ldots, n$ (repetitions allowed) showing exactly k increases between adjacent elements (Carlitz *et al.* (1966)).

4. Generating functions and factorial moments of the distribution

The probability generating function of the distribution (2.6), on using the polynomial (3.1), may be obtained as

$$(4.1)\qquad G_n(t;a,b,\theta) = \sum_{k=0}^{\infty} p_k(n;a,b,\theta)t^k = \frac{(-1)^n B_n(t;a,b,-1/\theta)\theta^n}{\binom{a+b+n-1}{n}},$$

and consequently the factorial moment generating function as

$$(4.2)\qquad F_n(t;a,b,\theta) = \sum_{r=0}^{n} \mu_{(r)}(n;a,b,\theta)\frac{t^r}{r!} = \frac{(-1)^n B_n(1+t;a,b,-1/\theta)\theta^n}{\binom{a+b+n-1}{r}},$$

where

$$\mu_{(r)}(n;a,b,\theta) = E[L_n(L_n-1)\cdots(L_n-r+1)], \qquad r = 1, 2, \ldots,$$
$$\mu_{(0)}(n;a,b,\theta) = 1.$$

A direct comparison of (4.2) to (3.13) yields

$$(4.3)\qquad \mu_{(r)}(n;a,b,\theta) = \frac{G(n,n-r;1/\theta,b/\theta)\theta^n}{\binom{a+b+n-1}{r}}, \qquad r = 1, \ldots, n.$$

Making use of the recurrence relation satisfied by the numbers $G(n,r,\alpha,\beta)$ (Charalambides and Koutras (1983)),

$$G(n,r;\alpha,\beta) = (\alpha r+\beta-n+1)G(n-1,r;\alpha,\beta) + \alpha G(n-1,r-1;\alpha,\beta),$$
$$r = 1, 2, \ldots, n, \quad n = 1, 2, \ldots,$$

or the recurrence (2.4), we can find a recursion formula for the factorial moments (4.3),

$$\mu_{(r)}(n;a,b,\theta) = \frac{r[n-r+b-(n-1)\theta]}{a+b+n-1}\mu_{(r-1)}(n-1;a,b,\theta) + \frac{a+b+n-r-1}{a+b+n-1}\mu_{(r)}(n-1;a,b,\theta).$$

For the first two factorial moments we obtain

$$\begin{aligned} &\mu_1(n;a,b,\theta) = E[L_n] = \frac{(1-\theta)\binom{n}{2}+bn}{a+b+n-1}, \\ &\mu_2(n;a,b,\theta) \\ &\quad = E[L_n^2]-E[L_n] \\ &\quad = \frac{2\left[3(1-\theta)^2\binom{n}{4}+(1-\theta)(1+3b-2\theta)\binom{n}{3}+b(b-\theta)\binom{n}{2}\right]}{(a+b+n-1)(a+b+n-2)}. \end{aligned} \tag{4.4}$$

Another point of interest is the following: employing the recurrence (3.3) one could show (for example by induction on n) that the polynomial $B_n(t;a,b,s)$ has n distinct real non-positive roots for all $n=1,2,\ldots$. As a consequence we deduce that

(a) $B_{n,k}(a,b,s)$ is a strictly logarithmic concave function of k, i.e.

$$[B_{n,k}(a,b,s)]^2 > B_{n,k+1}(a,b,s)B_{n,k-1}(a,b,s).$$

(b) The distribution (2.6) is unimodal either with a peak or with a plateau of two points (see Comtet (1974)).

(c) The random variable L_n can be expressed as a sum of n independent zero-one random variables.

5. Parameter estimation in the generalized Morisita model

Morisita (1971) proposed a heuristic method for estimating the environmental densities a and b. His method utilizes the ratio of the mean number of individuals in fine sand to the total number of individuals introduced i.e. $E[L_n]/n$. In order to get a second relation between the unknown parameters, Morisita estimated the ratio b/a through an additional series of experiments. Janardan (1988) gave the moment and maximum likelihood estimates based only on the initial experimental data.

In the present paper, we are going to estimate the environmental densities a, b and the repulsivity parameter θ by two different methods. Both of them are based on the initial Morisita's data only.

The first method proceeds along the lines set by Morisita. If a total of n individuals are allowed to choose their habitat, then, according to our generalized model, the mean number of individuals in fine sand is (see (4.4))

$$\frac{E[L_n]}{n} = \frac{(1-\theta)\frac{n-1}{2}+b}{a+b+n-1} = p_n.$$

Therefore

$$a = b\lambda - \frac{(n-1)(1-\lambda)}{2}(1-\theta) - (n-1)\theta, \qquad \lambda = \frac{1}{p_n}$$

and making use of Morisita's (1971) Table 2, we deduce that

$$a = 0.6313b - 2.6669\theta - 0.3331. \tag{5.1}$$

From the observed data of Table 1 (see also Table 2 of Morisita (1971) or Janardan (1988)) we may write

$$E[L_1] = \frac{b}{a+b} = \frac{29}{32}. \tag{5.2}$$

Finally, equating the probability that one out of 2 individuals selects fine sand to 13/32 (see Table 1 of Morisita (1971) or Janardan (1988)) we obtain

$$\frac{b-\theta}{a+b+1} \cdot \frac{b}{a+b} = \frac{13}{32}. \tag{5.3}$$

Equations (5.1), (5.2), (5.3) lead to a linear system in three unknowns which gives

$$a = 0.111, \qquad b = 1.076, \qquad \theta = 0.0956. \tag{5.4}$$

Next, we consider the minimum chi-square estimates of a, b, θ. If x_n denotes the observed number of ant lions (out of n) settled in fine sand, the chi-square statistic for the data of Table 1 is given by

Table 1. Observed and expected values for Morisita's experimental data.

n	Number of experiments	Number of ant lions introduced	Number of ant lions in fine sand: Observed x_n	Estimated (1)	Estimated (2)	Estimated (3)	Estimated (4)	Estimated (5)
1	32	32	29	28.3	24.5	24.4	29.0	29.1
2	32	64	45	42.6	40.5	40.6	44.7	43.6
3	32	96	58	58.1	56.5	56.7	59.7	58.6
4	30	120	67	69.2	67.9	68.1	69.7	69.2
5	29	145	81	81.3	80.2	80.4	80.6	80.7
6	22	132	71	72.6	71.8	72.0	71.2	71.7
7	10	70	39	38.0	37.6	37.7	36.9	37.3
χ^2 statistic				.2923	1.4696	1.4358	.2745	.2025

(1) Morisita, (2) Janardan (moment estimates), (3) Janardan (maximum likelihood estimates), (4) Heuristic estimates (5.4), (5) Minimum chi-square estimates (5.5).

$$\chi^2(a,b,\theta) = \sum_{n=1}^{7} \frac{(x_n - E[L_n])^2}{E[L_n]}.$$

Since $-a/n \leq \theta \leq b/n$, an exhaustive search for $a = 0$ to 1.5 with step 0.01, $b = 0$ to 1.5 with step 0.01 and $\theta = -a/7$ to $b/7$ with step 0.001 was performed and the following minimizing values of a, b, θ were obtained

$$a = 0.08, \qquad b = 0.81, \qquad \theta = 0.045 \tag{5.5}$$

yielding

$$\chi^2(0.08, 0.81, 0.045) = 0.2025.$$

Table 1 gives the observed and fitted values for Morisita's experimental data, for the five different available estimates of a, b, θ, i.e. Morisita's ($a = 0.086$, $b = 0.664$, $\theta = 0$), Janardan's moment estimates ($a = 0.2355$, $b = 0.7647$, $\theta = 0$), Janardan's maximum likelihood estimates ($a = 0.2449$, $b = 0.7928$, $\theta = 0$), our heuristic estimates ($a = 0.111$, $b = 1.076$, $\theta = 0.0956$) and minimum chi-square estimates ($a = 0.08$, $b = 0.81$, $\theta = 0.045$). The chi-square value is also provided for each case. The degrees of freedom for Morisita's and Janardan's models are 4, while for our model are 3 (since an additional estimate has been plugged in the chi-square statistic). Note that the introduction of the additional parameter θ in our model, leads to a substantial reduction of the chi-square error, especially compared to Janardan's estimates. As regards our heuristic estimates, their chi-square error is very close to the minimum chi-square error.

References

Carlitz, L. and Scoville, R. (1974). Generalized Eulerian numbers: combinatorial applications, *Journ. für die Reine und Angewandte Mathematik*, **265**, 110–137.

Carlitz, L., Roselle, D. P. and Scoville, R. A. (1966). Permutations and sequences with repetitions by number of increases, *J. Combin. Theory*, **1**, 350–374.

Charalambides, Ch. (1991). On a generalized Eulerian distribution, *Ann. Inst. Statist. Math.*, **43**, 197–206.

Charalambides, Ch. and Koutras, M. (1983). On the differences of the generalized factorials at an arbitrary point and their combinatorial applications, *Discrete Math.*, **47**, 183–201.

Comtet, L. (1974). *Advanced Combinatorics*, Reidel, Dordrecht, Holland.

Gould, H. W. and Hopper, A. T. (1962). Operational formulas connected with two generalizations of Hermite polynomials, *Duke Math. J.*, **29**, 51–63.

Janardan, K. G. (1988). Relationship between Morisita's model for estimating the environmental density and the generalized Eulerian numbers, *Ann. Inst. Statist. Math.*, **40**, 439–450.

Morisita, M. (1971). Measuring of habitat value by environmental density method, *Statistical Ecology* (eds. G. P. Patil *et al.*), 379–401, The Pennsylvania State University Press.