Spatial distribution of the montane unicorn

Stuart H. Hurlbert

Analysis of the spatial distribution patterns of five populations of the recently discovered montane unicorn (*Monoceros montanus*) yield several surprising results. Principal among these is the fact that when censused using a grid of 1 km² quadrats all populations yielded a variance:mean ratio of 1.0, though each population showed a different pattern of aggregation and none corresponded to a Poisson distribution. Distributions with these mathematical properties are termed *unicornian*. It is demonstrated that the variance:mean ratio is useless as a measure of departure from randomness, though it is widely recommended as such. The ratio is also uninterpretable as a measure of aggregation except for its coincidental relationship to indices such as the Morisita index of aggregation (\(I_M\)). \(I_M\) can be defined in terms of the probability of two randomly selected individuals being found in the same quadrat or sampling unit. It can be generalized to yield a family of indices, \(I_M^r\), relating to the probability of \(r\) randomly selected individuals being found in the same sampling unit. Presentation of a plot of \(I_M^r\) versus \(r\) may often be preferable to condensation of all information on aggregation into a single value, such as \(I_M\). Comment is also offered on the need to consider a range of spatial scales, on the corresponding definitional and scale aspects of plotless sampling, and on the fruitlessness of attempting to view spatial patterns themselves as consisting of only two distinct aspects, such as intensity and grain.

S. H. Hurlbert, Dept of Biology, San Diego State Univ., San Diego, CA 92182, USA.

"... the unicorn's horn has been so coveted that people have invented any number of untrue and malicious stories about the beast in order to justify its capture and killing. It is through this tangled maze of deceptions and half-truths that we must make our way, to try to get at the truth about ... unicorns... and to understand their unusual lives."

– Johnsgard and Johnsgard (1982)

For millenia, unicornology or, more properly, monocerology was mostly the private reserve of philosophers, poets and liars. Their accumulating extravagant claims completely overwhelmed the quieter, less numerous contributions of serious unicorn biologists. This eventually so tainted our perception of them that all national and international funding agencies ceased supporting scientific research on unicorns several centuries ago. And by the 18th century even the general public had abandoned belief in them (Shepard 1967). Nevertheless, a large amount of accurate information on the natural history of these beasts accumulated over the ages. Most of this has been summarized in a recent treatise by Johnsgard and Johnsgard (1982).

Four species of unicorn are known to science. Johnsgard and Johnsgard (1982) treat three of these: the karkadann (*Monoceros tyrannus* Johnsgard and Johnsgard) of India and Persia, now thought extinct; the ki-lin (*Monoceros orientalis* Johnsgard and Johnsgard) of eastern Asia; and the European unicorn (*Monoceros europus* Johnsgard and Johnsgard). The fourth species, the montane unicorn (*Monoceros montanus* della Roba) was not discovered until 1978 and has only recently been described (della Roba 1985). In the course of research on montane ecosystems during the last several years, I have obtained further information on this species. It has a cosmopolitan but very disjunct distribution. It now has been found on all the continents except Greenland and Antarctica but is restricted to mountainous regions. It is the only unicorn known from the New World. In the Old World it is everywhere allopatric or...
Fig. 1. Distributions of $q_k$ values for six unicorn (A-F) and five hypothetical (G-K) populations (solid lines). For comparison, the $q_k$ values representing the Poisson distribution (for $\mu = 10$, $Q = 1000$) are also shown in each figure (dashed lines). See Table 2 for provenances of populations.

allotopic with respect to the other unicorn species, except in Tibet where in the valleys in winter it occasionally occurs in mixed herds with the ki-lin.

Among the specific investigations I have undertaken on the montane unicorn is an analysis of spatial distribution of individuals within herds. Because the results on this investigation are so unusual, I dedicate this paper solely to them, reserving for future papers my findings on other aspects of their ecology.

The most striking finding is that within each population the unicorns have neither random nor regular nor aggregated distributions, the only three possibilities recognized by most scribes. Instead their distributions, though they vary much from one mountain region to another, are invariably, well, unicornian. No other organisms are known to exhibit them, and no other term exists for them.

These unicorn data provide the basis for a discussion of certain mathematical approaches to the analysis of spatial distribution. Specifically, I shall: 1) argue that the variance/mean ratio is of no value as a measure of aggregation, 2) argue that it is not reasonable to expect a single index to serve well both as a measure of departure from randomness (or the Poisson distribution) and
Tab. 1. Raw data \((q_k)\) values for unicorn and hypothetical populations. Labels correspond to those in Fig. 1 and Table 2. \(P = \) Poisson distribution with \(\mu = 10.0\).

<table>
<thead>
<tr>
<th>Population</th>
<th>(k)</th>
<th>(P)</th>
<th>(A)</th>
<th>(B)</th>
<th>(C)</th>
<th>(D)</th>
<th>(E)</th>
<th>(F)</th>
<th>(G)</th>
<th>(H)</th>
<th>(I)</th>
<th>(J)</th>
<th>(K)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Uniform</td>
<td>10</td>
<td>0.90</td>
<td>0.87</td>
<td>6992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. Extreme bimodal</td>
<td>10</td>
<td>203</td>
<td>2.93</td>
<td>1.00</td>
<td>197156</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I. Rectangular</td>
<td>10</td>
<td>29.7</td>
<td>1.20</td>
<td>0.37</td>
<td>5290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K. Maximum aggregation</td>
<td>10</td>
<td>99,900</td>
<td>999.9</td>
<td>1.00</td>
<td>359,289</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tab. 2. Statistical properties of the spatial distributions of six montane unicorn populations and five hypothetical populations. For all cases, \(Q = 1000\). Levels of statistical significance denoted as follows: ns, \(P > 0.05\); ***, \(P < 0.001\).

<table>
<thead>
<tr>
<th>Population</th>
<th>(u)</th>
<th>(I_M)</th>
<th>(D_P)</th>
<th>(X^2_{null})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicorn populations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Qomolungma</td>
<td>10</td>
<td>1.0*</td>
<td>0.27</td>
<td>493.2***</td>
</tr>
<tr>
<td>B. Matterhorn</td>
<td>10</td>
<td>1.0*</td>
<td>0.14</td>
<td>181.0***</td>
</tr>
<tr>
<td>C. Machu Picchu</td>
<td>10</td>
<td>1.0*</td>
<td>0.26</td>
<td>723.2***</td>
</tr>
<tr>
<td>D. Andean Altiplano</td>
<td>10</td>
<td>1.0*</td>
<td>0.13</td>
<td>91.4***</td>
</tr>
<tr>
<td>E. Grand Tetons</td>
<td>10</td>
<td>1.0*</td>
<td>0.13</td>
<td>103.7***</td>
</tr>
<tr>
<td>F. Torres del Paine</td>
<td>10</td>
<td>1.0*</td>
<td>0.72</td>
<td>3760 ***</td>
</tr>
</tbody>
</table>

Hypothetical populations

<table>
<thead>
<tr>
<th>Population</th>
<th>(u)</th>
<th>(I_M)</th>
<th>(D_P)</th>
<th>(X^2_{null})</th>
</tr>
</thead>
<tbody>
<tr>
<td>G. Moderate bimodal</td>
<td>10</td>
<td>41.7</td>
<td>1.32***</td>
<td>0.31</td>
</tr>
<tr>
<td>H. Extreme bimodal</td>
<td>10</td>
<td>203</td>
<td>2.93***</td>
<td>1.00</td>
</tr>
<tr>
<td>I. Rectangular</td>
<td>10</td>
<td>29.7</td>
<td>1.20***</td>
<td>0.37</td>
</tr>
<tr>
<td>J. Uniform</td>
<td>10</td>
<td>0</td>
<td>0.90***</td>
<td>0.87</td>
</tr>
<tr>
<td>K. Maximum aggregation</td>
<td>10</td>
<td>99,900</td>
<td>999.9***</td>
<td>1.00</td>
</tr>
</tbody>
</table>

a. To preclude excessively small expected \(q_k\) values, the \(q_k\) values for the left \((k<3)\) and right \((k>18)\) hand tails were pooled.

17* OIKOS 58:3 (1990)
as a measure of aggregation, 3) suggest a simple way of measuring departure from the Poisson distribution, 4) comment on the Morisita index of aggregation and 5) describe a generalized form of this index.

As defined above the focus of this paper is internationally narrow. It is concerned with the definition and quantification of aggregation in what are sometimes called “sparingly sampled patterns” (Diggle 1983). The data sets I present, however, represent fully sampled patterns. In using such, I seek to focus on the definitional or parametric form of aggregation indices. The separate albeit important problem of devising satisfactory sample estimators is discussed briefly but not tackled.

For fully sampled patterns (censused grids, mapped distributions, continuous transects), a greater variety of questions can be asked, and more sophisticated statistical and mathematical methodologies have been developed which make use of information on the spatial locations of sampling units. Spatial analysis, pattern analysis, and spectral analysis are among the labels used for various of these methods. Significant recent works on them include Goodall (1974), Fasham (1978), Ripley (1978), Greig-Smith (1979), Goodall and West (1979), Cox and Isham (1980), Cliff and Ord (1981), Carpenter and Chaney (1983), Diggle (1983), Ford and Renshaw (1984) and Renshaw and Ford (1984). These methods continue to be developed, refined and tested. Opinions as to what the various methods can and cannot accomplish are varied, and critical reviews of this literature would be useful. No attempt will be made here, however, to review these methods for fully sampled patterns.

A note on terminology: in the present context I regard the term aggregation, clumpedness, patchiness, and dispersion as essentially synonymous and generally will use only the first of these.

Field methods and calculations

The populations studied were from the following locations: the Grand Teton Range, Wyoming, U.S.A.; Machu Picchu in Peru; the Andean Altiplano of Bolivia; Torres del Paine in southern Chile; Qomolungma (Mt. Everest) on the China-Nepal border; and the Matterhorn on the Italy-Switzerland border.

Each unicorn herd consisted of exactly 10,000 individuals and occupied an area of exactly 1000 km². For censusing purposes each area was partitioned into 1000 quadrats of 1 km² each. The number of unicorns was determined for every quadrant. Since it is always as easy to count the unicorns in a thousand quadrats as in a single one, it seemed unprofitable to merely sample the areas. All censusing was done in the afternoon (1200–1800 hours local time) as previous research (Cole 1957) had shown that unicorns are most quiescent at that time.

The results of our censuses are given in Fig. 1 and Tables 1 and 2. For each population I have calculated the variance:mean ratio ($\sigma^2/\mu$), the Morisita index of aggregation ($I_M$), the new measure of departure from Poisson ($D_p$ described below), and the statistical significance of the departure of each $\sigma^2/\mu$ from 1.0 (with the statistic $\chi^2_{df=Q-1} = (Q-1)\sigma^2/\mu$ and of the departure of each observed distribution from the Poisson (by a $\chi^2$ goodness-of-fit test). These two statistical tests are, strictly speaking, superfluous or inappropriate for my data since all quadrats in each sampling universe were censused. They have been carried out simply to illustrate how different the results of the two tests can be.

To aid discussion I have also constructed five hypothetical distributions of individuals among quadrats (cases G, H, I, J, K) and applied these same indices and tests to them.

Variance:mean ratio

All unicorn populations examined had spatial distributions characterized by variance:mean ratios equal to 1.0. Given the widely differing locations of the six populations, it seems reasonable to suppose that a $\sigma^2/\mu = 1.0$ is a universal property of unicorn populations. To determine what such a property might imply about the spatial distribution and ecology of these populations, I examined the abundant literature.

The majority of works that describe the use of the variance:mean ratio as a method for analyzing pattern adopt the position of Greig-Smith (1964, 1983): “A measure of the degree of departure from Poisson expectation is therefore provided by the ratio of the variance to the mean [p. 59]... If the ratio of variance to mean is less than one, a regular distribution is indicated, if greater than one, a contagious distribution [p. 62 (63 in 1983 ed.)].” Similar statements can be found as far back as Svedberg (1922: 3) and Blackman (1942: 352) and in numerous recent books (e.g. Collier et al. 1973: 167, McLean and Ivimey-Cook 1973: 3467, McNaughton and Wolf 1973: 72, Poole 1974: 104, 112, Brower and Zar 1977: 121, Elliott 1977: 40, Southwood 1978: 27, 39, Cox 1980: 93, Sokal and Rohlf 1981: 87, Zar 1984: 410, Ludwig and Reynolds 1988: 27) and in a much greater number of articles.

All these authors imply that a $\sigma^2/\mu$ ratio exactly or approximately equal to 1.0 is evidence that a distribution conforms exactly or approximately to the Poisson distribution. This is indicated by their suggestions of $\sigma^2/\mu$ as a measure of “departure from Poisson” or “nonrandomness” and by their failure (except in the case of Greig-Smith) to recognize that the classification (regular, $\sigma^2/\mu < 1$; Poisson (random), $\sigma^2/\mu = 1$; aggregated (contagious), $\sigma^2/\mu > 1$) does not admit all possibilities. Only rarely (e.g. Stiteler and Patil 1972: 425; Pirolou 1974: 155) has the inadequacy of this classification scheme even been hinted at.

Many textbooks, not content to merely lead us down this primrose path, push us down it. Thus Odum (1971:
206) suggests that “If... the variance/mean ratio... is not
[significantly] different from 1, the distribution is ran-
dom.” Pianka (1978: 145) states: “When this ratio is
unity, the distribution of organisms in the quadrats fits
the Poisson distribution and the organisms are ran-
domly distributed with respect to the quadrats.” Van-
dermeer (1981: 131) states: “If the ratio is equal to one,
the pattern is random.” Shimwell (1971: 27), Smith
(1974: 734), Wetzel and Likens (1979: 335), and Brewer
and McCann (1982: 66) make similar claims.

Because so many authors have misstated the facts in
this case and because exercises on the analysis of spatial
distribution are so widely taught in introductory ecology
courses, it is probable that more misinformation has
been foisted upon students concerning the \( \sigma^2/\mu \) ratio
than concerning any other topic in ecology.

The Poisson distribution is, in fact, only one of an
infinite number of distributions for which \( \sigma^2/\mu = 1.0 \).
The unicorn populations demonstrate this very clearly.
Though all exhibit a \( \sigma^2/\mu \) of 1.0, all also deviate mark-
edly and significantly from the Poisson, as indicated by the \( \chi^2 \) goodness-of-fit tests.

How could so many be so wrong on such a simple
matter for such a long time? More than thirty years ago
Evans (1952) and Thomson (1952) pointed out the am-
biguous meaning of the \( \sigma^2/\mu \) ratio very clearly, Evans
even providing an example. Unfortunately, Evans
worked (and works) in Michigan, a very flat region
lacking in montane unicorns; and so his example of a
distinctly non-Poisson distribution with a \( \sigma^2/\mu = 1.0 \) had
to be a hypothetical one. Perhaps for that reason he was
took note of Evan’s (1952) example but made light of it,
saying: “Such a combination of generally regular distri-
bution with occasional groups of individuals is unlikely
in natural vegetation and would in any case be recog-
nized without reference to a test.” In thus dismissing
Evans’ (1952) objection, Greig-Smith seemed unaware
of how numerous and varied are the non-Poisson distri-
butions with \( \sigma^2/\mu = 1.0 \). Nor did he appreciate that
many of these will not be readily distinguished from the
Poisson by his ‘eyeball’ test. Because of the landmark
nature of his book, many modern ecologists may have
accepted Greig-Smith’s dictum on this topic without
critically analyzing it or going back to the older litera-
ture.

Of all recent writers on this matter, Pielou (1974: 155)
is perhaps clearest: “…one should not conclude that a
pattern is random because... the variance:mean ratio is
close to 1... This is an exceedingly common error, but
logically it is on a par with the following elementary
blunder: all crows are birds and therefore all birds are
crows.”

Clearly the \( \sigma^2/\mu \) ratio is not a good general measure of
“departure from Poisson.”

An index of departure from Poisson

It is an easy task to devise an index of the degree to
which a given distribution departs from a Poisson distri-
bution having the same mean, \( \mu \). The critical first step is
to discard the notion that such an index must also be
suitable as an index of aggregation. There is no index
that can fill both roles adequately.

The simplest index of departure from Poisson would
be based on the degree of overlap or concordance of the
observed distribution and the Poisson distribution of the
same \( \mu \). It is defined as:

\[
D_p = 1 - \sum_{k=1}^{w} \min (\pi_{kp}, \pi_{ka})
\]

(1)

\[
\pi_{kp} = \frac{q_{kp}}{Q}, \quad \pi_{ka} = \frac{q_{ka}}{Q},
\]

where

\( w \) = largest number of individuals observed in any
quadrat,
\( k = \) number of individuals per quadrat,
\( q_{kp} = \) number of quadrats containing \( k \) individuals, in
the observed or actual distribution,
\( q_{ka} = \) total number of individuals per quadrat,
\( q_{kp} = \frac{Q}{\mu^k e^{-\mu}} k! = \) number of quadrats expected to con-
tain \( k \) individuals, in a Poisson distribu-
tion with mean \( \mu \),
\( Q = \sum_{k=0}^{w} q_{ka} = \) total number of quadrats, in the sam-
pling universe or study site,
\( \mu = X/Q = \) mean number of individuals per quadrat,
\( X = \sum_{i=1}^{Q} x_i = \sum_{k=1}^{w} kq_{ka} = \) total number of individu-
als present in the \( Q \) quadrats, the sam-
ping universe, and
\( x_i = \) number of individuals in the \( i \)th quadrat

This index, \( D_p \), varies from 0 to 1, assuming the former
value when the observed distribution conforms exactly
to the Poisson, and the latter value (actually always
slightly less than 1) when a single quadrat contains all \( X \)
individuals present in the population. Geometrically it
represents 1.0 minus the area of intersection of the \( \pi_{kp} \)
and \( \pi_{ka} \) distributions with respect to \( k \).

It would be difficult to find a simpler, more assump-
tion-free measure of departure from Poisson than is \( D_p \).
On the other hand, it is clearly useless as an index of
aggregation. Though it will achieve its maximum value
when individuals are maximally aggregated, i.e. all in
one quadrat, there is no relationship between \( D_p \) and
aggregation (as measured by the Morisita index, \( I_M \),
for example; see Eq. 8) over most of the spectrum of pos-
sible distributions. \( D_p \) can vary widely without any ac-
companying change in \( I_M \): Note that our unicorn pop-
ulations have $D_p$ values ranging from 0.13 to 0.72, though all these populations have an $I_m = 1.0$ (Table 1). And $I_m$ can vary widely without any change in $D_p$, as shown by the data for hypothetical populations H and K.

As further evidence of the unrelatedness of $D_p$ and aggregation, one may note that uniform distributions (every quadrat with the same number of individuals) yield $D_p$ values intermediate to those for random and those for highly aggregated distributions. For the special case where $\mu$ is an integer, $D_p$ for a uniform distribution is exactly equal to one minus the $(\mu + 1)$th term in the Poisson series, i.e.

$$D_p (\text{uniform distribution, } \mu \text{ an integer}) = 1 - (\mu e^{-\mu/\mu}) \tag{2}$$

Thus for uniform distributions with $\mu = 1, 2, 5$ and 10, $D_p$ values are 0.63, 0.73, 0.82, and 0.87, respectively. For Poisson distribution with $\mu = 1, 2, 5$ and 10, $D_p \equiv 0$; this approximation increases with increasing $Q$, simply because the degree to which an actual distribution can approximate the Poisson increases with $Q$. For maximally aggregated distributions, $D_p$ is equal to 1.0 minus the $(X+1)$th term in the Poisson series, i.e.

$$D_{\text{max}} = 1 - (\mu^X e^{-\mu/X}) \tag{3} = 1, \text{ for all but trivial cases (e.g. } X=2, Q=2)$$

An attractive feature of $D_p$ is that the conventional $\chi^2$ goodness-of-fit test is, in fact, a test of the null hypothesis that $D_p = 0$. Of course, if we have censused all the quadrats, then a goodness-of-fit test or any other significance test usually is irrelevant. One exception would be if we chose to regard as the population of interest not the particular grid of censused quadrats but all possible realizations of some random mechanism imagined to generate the spatial patterns.

$\sigma^2/\mu$ as index of aggregation

Though a $\sigma^2/\mu$ ratio of 1.0 is not unique to the Poisson distribution and therefore not suitable as a measure of “departure from Poisson”, this does not automatically disqualify the $\sigma^2/\mu$ ratio as a measure of aggregation or dispersion. We may decide that it is proper that all distributions, Poisson and otherwise, having a $\sigma^2/\mu$ of 1.0 be deemed equally aggregated. But are there rational grounds for using the $\sigma^2/\mu$ ratio as an index of aggregation?

The $\sigma^2/\mu$ ratio does meet one universally agreed upon requirement for an index of aggregation: it has a minimum value when every quadrat contains the same number of individuals and a maximum value when all individuals are in a single quadrat. However, a potentially infinite number of other indices also have this property. Mathematically the critical question is: exactly how do we want an index of aggregation to behave between these two extreme situations? The answer should be specified by our decision, ideally reached on ecological grounds if we are dealing with ecological data, as to how aggregation should be defined. No one has expressed this more clearly than Pielou (1977: 124): “...the phrase ‘degree of aggregation’ describes a vague, undefined notion that is open to several interpretations. If aggregation is to be measured we must first choose from a number of possibilities some measurable property of a spatial pattern that is to be called its aggregation, and the method of measurement is then implicit in the chosen definition. Thus the several existing ways of measuring aggregation are not different methods of measuring the same thing: they measure different things.” Taylor (1980: 1331) is also cogent on the point: “there can be no statistical definition of [aggregation] until it is biologically defined.”

As a measure of aggregation the $\sigma^2/\mu$ ratio must fail because it lacks clear interpretability. If $\sigma^2/\mu = 1.5$, what does this mean? There is no reply that is not circular. One might say it means that aggregation is 50% greater than it would be if the individuals were distributed at random. But, says the critic, exactly what do you mean by “aggregation?” The honest but circular reply has to be – “I mean the property measured by the $\sigma^2/\mu$ ratio.” The use of the $\sigma^2/\mu$ ratio as an index of aggregation has not been based on an explicit definition of aggregation. It has been selected and used only on the basis of two faulty premises: (1) that the $\sigma^2/\mu$ ratio is a reasonable measure of “departure from Poisson” (whereas in fact it will not detect many kinds of even extreme departures from the Poisson, as already shown); and (2) that a good measure of departure from Poisson should be a good measure of aggregation (whereas in the absence of a prior, specific definition of aggregation there is no reason to assume that a given Poisson distribution cannot exhibit the same degree of aggregation as a given non-Poisson distribution).

Previous objections to the $\sigma^2/\mu$ ratio as a measure of aggregation usually have been based not on the absence of an a priori definition of aggregation in biological terms but rather on the mathematical properties of the ratio (e.g. Morisita 1959, 1962, Cole 1946, Green 1966, Lefkovitch 1966, Lloyd 1967). With Pielou (1977), I feel these to be only of secondary importance. If we have defined aggregation in a clear and considered way, the equation to be employed in its calculation will be specified more-or-less automatically. If we dislike the mathematical properties of the index, then we have the option of reconsidering our conceptual of aggregation. If we revise this, the revision will automatically lead to a new index. Yet the more usual approach has been to tinker with an index until it assumes the desired mathematical properties, all the while ignoring the fact that every modification implicitly redefines “aggregation”. Statisticians are especially prone to producing lengthy works on this topic that never confront the problem of definition in a considered a priori manner.
The Morisita and Lloyd indices

For quadrat data, only two indices interpretable in straightforward and non-circular terms has been put forward, those of Morisita (1959) and Lloyd (1967).

If two individuals are selected (without replacement) at random from a population of X individuals distributed over an area that has been completely partitioned into Q quadrats, the probability that both individuals will be in the same quadrat is

\[ P = \sum_{i=1}^{Q} \left( \frac{x_i}{X} \right) \left( \frac{x_i-1}{X-1} \right) \]

(4)

\[ = \sum_{k=2}^{w} \left( k \right) \left( \frac{k-1}{X-1} \right) \]

(5)

Now suppose that this same population of X individuals were to be distributed at random, so that the number of individuals per quadrat followed a multinominal distribution. In this case, \( \Delta_s \) would have an expectation, \( \Delta_p \),

\[ \Delta_p = \frac{1}{Q} \]

(6)

The ratio of these two probabilities makes an attractive index of aggregation, namely, in parametric form,

\[ I_M = \frac{\Delta_s}{\Delta_p} = Q \sum_{i=1}^{Q} \left( \frac{x_i}{X} \right) \left( \frac{x_i-1}{X-1} \right) \]

\[ = \left( \frac{X}{X-1} \right) \left( \frac{1}{\mu} \right) \left( \frac{\sigma^2}{\mu} + \mu - 1 \right) \]

(7)

(8)

(Morisita 1959, 1964, 1971)

\( I_M \) measures how many times more likely it is that two randomly selected individuals will be from the same quadrat than it would be if the X individuals in the population were distributed at random. For example, if \( I_M = 1.5 \) then the probability that the two randomly selected individuals are from the same quadrat is 50% greater than it would be in the case of a random distribution.

Several years after Morisita (1959) first proposed this “index of dispersion”, Lloyd (1967) independently derived an index which he called “patchiness”. As he defined it, patchiness (\( I_L \)) measures “how many times as ‘crowded’ an individual is, on average, as it would... be if the same population had a random distribution”, where the crowding each individual experiences is measured as the number of other individuals occurring in the same quadrat that it occupies. In parametric form,

\[ I_L = \left( \frac{1}{\mu} \right) \left( \frac{\sigma^2}{\mu} + \mu - 1 \right) \]

(9)

For most purposes \( I_M \) and \( I_L \) may be regarded as identical. In the remainder of this paper I usually shall refer explicitly only to \( I_M \) but it should be understood that, in the absence of caveats to the contrary, all statements apply to \( I_L \) as well. Certainly if X is very low (e.g. 10) \( I_M \) will be noticeably higher (e.g. by 11%) than \( I_L \). Even such differences, however, usually will be negligible given the range of values (zero to Q) assumable by \( I_M \) or \( I_L \) and given that sample estimates of either index for a given population also can range over orders of magnitude when aggregation is high (Lloyd 1967: 13).

At least initially, Morisita’s contribution to the measurement of aggregation was less appreciated by ecologists than was Lloyd’s. I suspect one reason for this was that in his early papers Morisita (1959, 1962) was never explicit about the simplicity and appropriateness of the definition of aggregation implied by \( I_M \); instead he emphasized its mathematical properties, though without clearly distinguishing the parametric and sample estimator forms of \( I_M \) (Patil and Stiteler 1974). Only later did he point out how it might be interpreted in terms of probability of co-occurrence or crowding (Morisita 1964, 1971). Another reason may be that both Morisita (1959: 220) himself and a few other authors (Poole 1974: 117, Stiteler and Patil 1971: 444–446) have implied that use of \( I_M \) as an index of aggregation or patchiness should be contingent on certain restrictive assumptions being met. The assumptions in question need be invoked, however, only when \( I_M \) is to be used as more than a simple descriptive index – for example, when one wishes to use it in pattern analysis for such dubious (Pielou 1977: 141, Vandermeer 1981: 139) enterprises as the detection of “clump size”.

It is apparent that \( I_M \) is a simple function of the \( \sigma^2/\mu \) ratio and that \( I_M \) defines all distributions, Poisson or unicornian, having \( \sigma^2/\mu = 1.0 \) as being equally aggregated. Through this relation an ecological interpretation of the \( \sigma^2/\mu \) ratio was made apparent for the first time (Lloyd 1967). Nevertheless, this in no way justifies the use of the \( \sigma^2/\mu \) ratio itself as an index of aggregation; by itself it cannot be interpreted except in a circular way, as indicated earlier.

Since \( I_M \) can deviate from a value of 1.0 only to the extent that \( \sigma^2/\mu \) also deviates from a value of 1.0, we can test whether the observed degree of aggregation is significantly or less than a randomly distributed population would exhibit. We do this by using the test criterion \( \chi^2 = (Q-1)s^2/m \) (David and Moore 1954), where \( Q \) is the number of quadrats in a sample and \( s^2 \) and \( m \) are the sample estimates of \( \sigma^2 \) and \( \mu \). However, it is misleading to state (Morisita 1962: 2) that this represents “a test for the significance of departure from randomness”. It represents only a test of the significance of a departure of \( s^2/m \) from 1.0.

\( I_M \) versus other indices

I wish now to present a generalized version of \( I_M \). Before doing so, however, it is necessary to correct state-
ments made about the properties of $I_M$ and other indices in three recent reviews of this topic (Taylor 1984, Ludwig and Reynolds 1988, Ch. 3, Krebs 1989, Ch. 4).

Krebs (1989: 166) concluded that “For quadrat sampling the two best [indices of dispersion] seem to be Green’s [$C_x$] and the standardized Morisita coefficient [$I_p$, of Smith-Gill], both of which are relatively unaffected by population density [$\mu$] and sample size [Q or $Q_i$]”. He cites a comparative analysis by Myers (1978) as a foundation for this conclusion. Myers, however, did not claim that these indices were unaffected by sample size, though Elliott (1977: 74) did make that claim for $C_x$.

Krebs also stated (p. 150) that $I_M$ “is relatively independent of population density but is affected by sample size”. The claim concerning sample size dependence seems to be based on too uncritical an acceptance of Green’s (1964: 4) statement that $I_M$ “is a strong function of [sample size] at high positive and negative contagion [and therefore] a good index of non-randomness [sic] [only when sample size is the same] for all sets of samples to be compared.”

This is indeed a pot of porridge. In fact, both $C_x$ and $I_p$ are strongly influenced by sample size while $I_M$ is not, just the reverse of Krebs’ claims. For example, if a population of quadrats with $\mu = 2.0$ and $\sigma^2 = 2.8$ is randomly sampled with sample sizes ($Q$) of 10 and 100, respectively, the respective calculated values will be about: 0.02 and 0.002 for $C_x$; 0.10 and 0.50 for $I_p$ (for $\alpha = 0.05$); and 1.2 and 1.2 for $I_M$. The qualifier “about” simply acknowledges the likelihood of sampling error.

The strong dependence of $C_x$ and $I_p$ on sample size ($Q$) is one of the most negative features of these indices; and the independence of $I_M$ in this regard is of its several attractive features. The matter of dependence on mean density ($\mu$) is a slightly more complex matter. Before considering that, let us dispose of $I_p$.

The “standardized Morisita index”, $I_p$, of Smith-Gill (1975) is most misleadingly so labelled. It is essentially a test statistic. It is not only a function of sample size, it is also a function of $\alpha$, the level of significance, which must be specified before $I_p$ can be calculated. The lower the value of $\alpha$, the lower the absolute value of $I_p$. $I_p$ is so defined that only when $|I_p| \geq 0.5$, will $I_p$ be significantly different from zero. Values of $I_p$ between $-0.5$ and $+0.5$ can never be statistically significant no matter how large the samples from which they are calculated. Its bizarre derivation makes $I_p$ uninterpretable in terms of crowding or encounters or co-occurrences or other biologically relevant phenomena. Despite the strong recommendations of Myers (1978) and Krebs (1989), $I_p$ is egregiously unsuitable as a measure of aggregation.

One can more logically devise a “standardized Morisita coefficient” by dividing $I_M - 1$ by the lowest (if $I_M \leq 1$) or the highest (if $I_M > 1$) value of $I_M - 1$ possible for the given values of $X$ and $Q$. This index, $I_{MS3}$, will equal $-1$ for a maximally uniform distribution of organisms over quadrats, 0 for a random distribution, and 1 for a maximally aggregated distribution.

The simplicity and familiarity of this finite range of values is beguiling to many. Also, some authors (e.g. Green 1966, Smith-Gill 1975, Myers 1975, Krebs 1989) clearly would feel it appropriate that $I_{MS3} = 1$ when all $X$ individuals are in a single one of the $Q$ quadrats regardless of the values of $X$ and $Q$. Yet a consequence of such “standardization” is that $I_{MS3}$ is strongly influenced by sample size when $I_{MS3} > 0$ and by mean density when $I_{MS3} < 0$.

Most discussions of the dependence of aggregation indices on population density are confused. At least four different criteria for assessing dependence have been used. Only one of these seems useful (“effect of random death”). One seems unimportant (“value of $I_{max}$”) and the third and fourth are simply inappropriate (results of simulation studies like that of Myers, 1978, and of tests with “real biological populations” like those of Taylor, 1984). Critical evaluation of these follows.

The only reasonable general criterion of density dependence that has been put forward is the behavior of an index when “in a given population, a proportion of the individuals [is] selected at random and killed or removed” (Piérou 1969: 92). Piérou showed that $I_M$ would remain unchanged by such random mortality “whatever the form of the parent distribution”. Hill (1973) states such density independence will characterize any index of the form $(\sigma^2-\mu)/\mu^2$. In this most fundamental way, $I_M$ is independent of population density whereas most other indices, including $C_x$ and $I_p$, are not, contrary to statements by Elliott (1977), Myers (1978), and Krebs (1989).

Though Ludwig and Reynolds (1988) acknowledge that $I_M$ and $I_p$ are independent of $X$ or $\mu$ by the random removal criterion, they discuss both as indices of “pattern intensity” in contradistinction to “indices of dispersion”, such as $C_x$ and the variance :mean ratio, which are treated in a separate section. The distinction is unwarranted. But it and their belief that $C_x$ is independent of $X$ or $\mu$ explain why they regard $C_x$ as the “index of clumping” that is “most recommendable”.

How, under conditions of maximum aggregation, density affects the value of aggregation indices has been the aspect of density dependence of primary concern to many workers. This criterion, on analysis, seems of minor importance, however. The condition of maximum aggregation is vanishingly rare. It thus will be equally rare that the observed values of indices such as $I_M$ will be limited by density ($\mu$, or total number of individuals, $X$). On the other hand, indices which are designed to always yield the same value under conditions of maximum aggregation (e.g. $C_x$ and $I_{MS3}$) invariably are strongly dependent on density under all other conditions.

Exercises of the sort utilized by Myers (1978) to assess the density dependence of aggregation indices simply are not valid for that purpose. One cannot accept
her conclusion, concurred in by Krebs (1989), that "[C\textit{r}] and [l\textit{ij}] are not correlated to mean density and are the best candidates for use when analyzing actual changes in distribution of organisms with changes in density... [and] there was a decrease in [IM\textit{r}] with increased density as would be expected as a statistical artefact". Those statements misrepresent the actual properties of these indices as described earlier in this section and as they are evident from the definitional formulas of the indices.

Myers' procedure did not test the dependence of aggregation indices on density but only tested, under a specific set of conditions, whether various aggregation indices tended to be correlated with the particular aggregation index (the "clumping variable") she used to create the pattern in her simulated populations. While she herself did not view her clumping variable as an index of aggregation, it was in effect just that. Her analyses showed that when indices were calculated for populations having different densities but the same "clumping variable", some indices (e.g. C\textit{r}, I\textit{p}) showed no strong tendency to vary with "density" (i.e. to be invariate like the clumping variable itself), while others (e.g. I\textit{M}) were strongly correlated with "density". Interpretation of these results apparently was based on the mistaken assumption that her simulated populations that were invariate with respect to the "clumping variable" would also be invariate with respect to any other aspects of pattern that might contribute to aggregation as that is defined by the various indices.

(Myers' analyses gave slightly different results for I\textit{M} and I\textit{r}. Probably this was a consequence not of the difference between these two indices but rather of her fitting her data to negative binomial distributions and calculating I\textit{r} as 1 + (1/k) instead of directly from the x\textit{i} values, as she did in the case of I\textit{M}. If, as is probable, the simulated distributions did not conform to the negative binomial, 1 + (1/k) would have been a biased estimator of I\textit{M} (Pielou 1969: 96.).

Taylor (1984) makes a number of incorrect statements about I\textit{M} and, not surprisingly, concludes the index is not "successful". He states that I\textit{M} was developed "initially to investigate the effects of quadrat size on the average number of individuals" (p. 335), and that I\textit{M} "was devised to remove the effects of quadrat size on the measure of aggregation" (p. 336). The former objective would seem trivial and the latter impossible, but in any case they were not objectives of either Morisita (1959, 1962) or Lloyd (1967).

Taylor implies (p. 337) that I\textit{M} = 1 + (1/k), whereas this holds only when the observed q\textit{k} distribution closely conforms to a negative binomial distribution (see above). He states that I\textit{M} is dependent on mean density (p. 337, 340) as are "all measures of aggregation" (p. 342). But at least by the "random death" criterion this claim is also incorrect. It reflects the general confusion in his review between the mathematical dependence (or lack thereof) of dispersion indices on density and the correlations that may exist, as a consequence of biological phenomena, between degree of aggregation and density in real populations. These latter correlations may contain some interesting biology, but they are an unreliable means for assessing the desirability, appropriateness, or intrinsic properties of a given index.

In sum, the criticisms of I\textit{M} by Myers (1978), Taylor (1984) and Krebs (1989) and the favorable appraisals of C\textit{r} and I\textit{p} by the first two authors and Ludwig and Reynolds (1988) are all unwarranted.

The I\textit{Mr} family of indices

It would be very convenient if a simple index such as I\textit{M} captured all the information on the dispersion patterns in a population. It cannot do this of course. In fact, if we insist on all the information, we are calling for a map showing the exact position of each individual. The following represents a compromise approach where additional information on dispersion patterns can be extracted without excessive complexity or additional assumptions.

I\textit{M} (Eq. 8) is based on the idea of selecting two individuals at random. However, by considering the drawing of any number, r, of individuals at random, we can define a family of indices, I\textit{Mr}, r = 2, 3, 4, ..., w. As r increases, I\textit{Mr} becomes dependent only on those quadrats showing the higher degrees (k \geq r) of aggregation or crowding.

We begin by noting that if r individuals are selected at random (without replacement) from an actual population of X individuals, the probability that all r individuals are in the same quadrat is

\[ \Delta_{se} = \sum_{i=1}^{X} \frac{X_i}{X} \cdot \frac{X_{i-1}}{X-1} \cdots \frac{X_{i-r+1}}{X-r+1} \]  
\[ = \sum_{k=r}^{w} q_k \left( \frac{k}{X} \right) \cdot \frac{k-1}{X-1} \cdots \frac{k-r+1}{X-r+1} \]  
\[ = \frac{(X-r)!}{X!} \cdot \sum_{k=r}^{w} q_k \cdot k! \cdot \frac{(k-r)!}{(k-r)!} \]

the generalized form of Eq. 8.

Similarly, if the X individuals are randomly distributed among quadrats, the probability that r randomly selected individuals would be found in the same quadrat is given as

\[ \Delta_{pr} = 1/Q^{r-1} \]  

The generalized form of I\textit{M} therefore is defined as

\[ I_{Mr} = \Delta_{sr}/\Delta_{pr} \]
fewer than \( r \) individuals per quadrat (so long as \( X_k < r \) and \( Q_k > r \) remain unchanged). For example, if \( r = 3 \) then two \( q_k \) distributions that are identical for \( k \leq 3 \) will yield identical \( I_{M3} \) values even though their \( q_k \) distributions for \( k < 3 \) may be very dissimilar, e.g. \( \{ q_0 = 5, q_1 = 0, q_2 = 5 \} \) versus \( \{ q_0 = 0, q_1 = 10, q_2 = 0 \} \). Only \( I_{M2} \) (Eq. 8) is based on the information in the entire \( q_k \) distribution.

It is apparent from the above property of \( I_{Mr} \) that it defines, for \( r > 2 \), aggregation as a threshold phenomenon. A given individual in the \( r \)th quadrat shares its quadrat with \( (x_i - 1) \) other individuals, but only \( (x_i - r + 1) \) of these cause crowding or contribute to aggregation as defined and measured by \( I_{Mr} \). For example, if \( r = 5 \), then the intensity of crowding is the same in all \( Q_k < r \) quadrats where \( k < r \), that is, for all individuals accompanied by \((=\) sharing a quadrat with\) 0, 1, 2, or 3 other individuals. There probably are only rare situations where this mathematical property of \( I_{Mr} \) corresponds to a real biological phenomenon, such as individuals being physiologically or behaviorally unaffected by other individuals except where density (= number per sampling unit) exceeds some sharply defined critical value.

Patil and Stiteler (1974: 248) also developed a generalized extension of Lloyd's (1967) patchiness index. Their index, \( R_r \), has a more complex interpretation than does \( I_{Mr} \), however, and seems even less likely to correspond to any real biological phenomena than does \( I_{Mr} \). \( R_r \) is based on “mean crowding” being defined as “the mean number per group of size \( r \) of other individuals in the same quadrat” (Patil and Stiteler 1974).

### Behavior of \( I_{Mr} \) curves

An \( I_M \) value measures only the average degree of excess crowding relative to that characterizing a Poisson distribution. The \( I_{Mr} \) curves reflect both that value and the particular pattern, out of many possible, of \( \Delta q_k (= q_{ka} - q_{kp}) \) values that gave rise to it. For example, did all individuals experience the same degree of crowding?

![Fig. 2. \( I_{Mr} \) curves for six unicorn (A-F) and five hypothetical (G-K) populations.](Image)

\[
I_{Mr} = \frac{Q_r^{-1} (X-r)!}{X!} \sum_{k=r}^{w} q_k \cdot k! \cdot k^{-1} (k-r)! \tag{14}
\]

It measures the degree to which the probability of finding all \( r \) individuals in the same quadrat is greater or less than it would be if individuals were randomly distributed. It equals 1.0 for all \( r \) for random distributions and for non-random distributions can assume values from 0 to \( +\infty \).

For a given distribution, \( I_{Mr} \) may be calculated for all values of \( r \) from 2 to \( w \) and the results plotted on semi-log graph paper. Alternatively, we may wish to plot a graph of \( r \) versus the actual probabilities, \( \Delta_{ap} \) and \( \Delta_{ac} \). Figs 2 and 3 shows the first approach as applied to the unicorn populations and some hypothetical ones. Note that for a given population \( I_{Mr} \) values may all be >1.0 (dropping to zero only at \( r = w + 1 \)), or all <1.0, or some may be <1.0 and some >1.0.

Each \( I_{Mr} \) value that can be calculated for a distribution measures a different aspect of aggregation. As \( r \) increases, only the aggregation that occurs in quadrats where \( k \geq r \) contributes to the value of \( I_{Mr} \). More specifically, the value of \( I_{Mr} \) is completely uninfluenced by the manner in which the \( X_k \) individuals are distributed among the \( Q_k \) quadrats containing fewer than \( r \) individuals per quadrat (so long as \( X_k < r \) and \( Q_k > r \) remain unchanged). For example, if \( r = 3 \) then two \( q_k \) distributions that are identical for \( k \leq 3 \) will yield identical \( I_{M3} \) values even though their \( q_k \) distributions for \( k < 3 \) may be very dissimilar, e.g. \( \{ q_0 = 5, q_1 = 0, q_2 = 5 \} \) versus \( \{ q_0 = 0, q_1 = 10, q_2 = 0 \} \). Only \( I_{M2} \) (Eq. 8) is based on the information in the entire \( q_k \) distribution.

It is apparent from the above property of \( I_{Mr} \) that it defines, for \( r > 2 \), aggregation as a threshold phenomenon. A given individual in the \( r \)th quadrat shares its quadrat with \( (x_i - 1) \) other individuals, but only \( (x_i - r + 1) \) of these cause crowding or contribute to aggregation as defined and measured by \( I_{Mr} \). For example, if \( r = 5 \), then the intensity of crowding is the same in all \( Q_k < r \) quadrats where \( k < r \), that is, for all individuals accompanied by \((=\) sharing a quadrat with\) 0, 1, 2, or 3 other individuals. There probably are only rare situations where this mathematical property of \( I_{Mr} \) corresponds to a real biological phenomenon, such as individuals being physiologically or behaviorally unaffected by other individuals except where density (= number per sampling unit) exceeds some sharply defined critical value.

Patil and Stiteler (1974: 248) also developed a generalized extension of Lloyd's (1967) patchiness index. Their index, \( R_r \), has a more complex interpretation than does \( I_{Mr} \), however, and seems even less likely to correspond to any real biological phenomena than does \( I_{Mr} \). \( R_r \) is based on “mean crowding” being defined as “the mean number per group of size \( r \) of other individuals in the same quadrat” (Patil and Stiteler 1974).

### Behavior of \( I_{Mr} \) curves

An \( I_M \) value measures only the average degree of excess crowding relative to that characterizing a Poisson distribution. The \( I_{Mr} \) curves reflect both that value and the particular pattern, out of many possible, of \( \Delta q_k (= q_{ka} - q_{kp}) \) values that gave rise to it. For example, did all individuals experience the same degree of crowd-
ing or did some experience much higher degrees of crowding than did others?

An $I_M$ curve and a $\Delta q_k$ distribution are just two different ways of representing the same information. They are analogous in this regard to the species richness curves (Sanders 1968, Hurlbert 1971) and rank-abundance curves (Whittaker 1972) that can be used to present information on species abundance relations or taxonomic structure. Just as those two approaches, employed jointly, are more informative than the mere calculation of diversity indices (James and Rathbun 1981), so the joint presentation of $I_M$ curves and $\Delta q_k$ distributions seems likely to lead to clearer and less facile interpretations of spatial distribution patterns than will reliance on $I_M$ values alone.

For $\mu$ fixed at 10.0, the curve for maximum aggregation (K) and that for a uniform distribution (J) define the domain within which all other $I_M$ curves must fall. This is not much of a constraint. For example non-zero values for $I_M$ range over about three orders of magnitude, those for $I_M^2$ over about 23, and those for $I_M^3$ over about 50. The greater range of $I_M$ values possible at higher $r$ is a consequence of their increasing dependence on only the right hand ($k>r$) portion of the $q_k$ distributions. The means for this part of the reference Poisson distribution and this part of any of the other distributions can become very discrepant as $r$ increases, and this permits greater discrepancies in $I_M$ values than are possible when means are the same.

Comparison of the $I_M$ curves for the hypothetical distributions (H-K) shows that increasing the thickness and/or length of the right- and left-hand tails of the $q_k$ distribution tends to increase the $\sigma^2$, $I_M$, the maximum value attained by $I_M$, and the $r$ value at which this maximum $I_M$ value is attained (Table 2, Fig. 2). "Tails" is used broadly here to refer to distinct modes centered on $k=30$ in Fig. 2H and at $k=0$ in Fig. 2K as well as tails of the more properly tapering sort.

The $I_M$ curves for the unicorn populations (A-G) occupy only a small portion of the domain bounded by $I_M$ curves for the hypothetical K and J populations. Though by definition all unicorn populations had the same $I_M$ value, $I_M$ values for larger $r$ values ranged over several orders of magnitude. The greater the thickness and/or length of the right hand tail of the $q_k$ distribution, the more the $I_M$ curve shifts upward and to the right. Because the $\sigma^2$ is fixed at 10.0, lengthening of the right-hand tail must be accompanied by a foreshortening of the left-hand tail and/or an increase in $q_k$ values for $k$ values equal to or somewhat less than $\mu$ (e.g. Fig. 2A).

So long as it is sufficient to maintain $\mu = \sigma^2 = 10.0$, the particular shape of the left-hand portion of the $q_k$ distribution seems to have little influence on the unicornian $I_M$ curves. Thus the Machu Picchu (C) and Torres del Paine (F) unicorn populations both have $q_k$ distributions sharply truncated at $k=14$ (Fig. 1) and very similar $I_M$ curves (Fig. 2). Yet they differ radically in the left-hand portions of their $q_k$ distributions.

Ultimately all $I_M$ curves must curve downwards, i.e. develop a negative slope. This is a consequence of the fact that the Poisson distribution extends, in theory, to infinite values of $k$ whereas the hypothetical, unicorn, and other observed distributions do not. Even the curve for population K eventually turns downward, though we would need another meter of page length to show this!

Each $I_M$ curve shown terminates at $r=w$. For all $r>w$, $I_M$ is defined, but it equals zero and is not conveniently represented on the logarithmically scaled ordinate.

The occasional intersection of $I_M$ curves (e.g. G and H; I with A and B; C and F – see Figs 2, 3) shows how the relative patchiness or aggregation of two populations can change when only higher degrees of crowding are allowed to influence the index of aggregation. For example, $I_M$ is 20% greater for population I than for population A, but $I_M^2$ is almost 1000% greater for population A than for population I. In general $I_M$ may remain the best single index of aggregation because it is based on the entire $q_k$ distribution. But if biological effects of crowding in, say, populations A and I occur only where the number of individuals per quadrat exceeds some threshold, our understanding of the biology will be enhanced by inspection of the entire $I_M$ curves. If we found evidence of greater biological stress in population A than in population I, consideration of only $I_M$ values would not lead us to consider differential crowding as an explanation. Inspection of the $I_M$ curves would suggest that crowding could be a very good explanation if there were independent evidence that the biological effects of crowding only occurred at densities greater than, say, 10-12 individuals per quadrat.

**Sample estimators**

In practice, investigators often select a certain number of quadrats or sampling units at random from a larger sampling universe of quadrats. In these situations, it is desirable to have available an unbiased sample estimator (I) of the parameter or index (I) of interest and an estimate of the standard error associated with it.

Some sample estimators have been proposed that are of interest here. Morisita (1971: 3) suggests as the sample estimator of $I_M$

$$I_M = \frac{s^2 - m}{m^2 - (s^2/Q_s)} + 1 \quad (15)$$

Lloyd (1967) offered another approach based on the assumption that the $q_k$ values followed a negative binomial distribution. Reed (1983) uses a jack-knife method and obtains sample estimates and confidence intervals very close to those obtained by Lloyd (1967). Hutchinson and Lyons (1989) derive the moments of $I_M$ for the case where the values follow the negative binomial and Poisson distributions and claim that their approach less frequently leads to underestimation of the variance than does Reed’s approach. I will not attempt here to eval-
uate or extend any of these approaches to the estimation problem. It is worth noting, however, that as \( r \) increases the error associated with \( I_{Mr} \) must increase because the \( I_{Mr} \) values will be based on an increasingly small subset \( \left( \sum_{k=r}^{\infty} q_k \right) \) of the total number of quadrats in the sample. Clearly the right-hand portions of \( I_{Mr} \) curves must be interpreted with caution.

**Interpretation of unicorn spatial patterns**

I beg the reader’s pardon for such a lengthy excursion into methodology when it was the unicorns themselves that were to be the focus of this paper. The following is a summary of what may be inferred from our data about the ecology of these beasts.

The complete lack of differences among unicorn populations in size (\( X = 10,000 \) individuals), area occupied (\( A = 1000 \text{ km}^2 \)), and equality of mean and variance (\( \mu = \sigma^2 = 10.0 \)) perhaps represent divinely imposed properties. At least the laws of genetics, ecology, and statistics are incapable of accounting for this invariance as a natural phenomenon. Nor do they provide clues as to the significance of the mean values of these parameters.

One of the most striking characteristics of the populations is that each has a unique, non-Poisson \( q_k \) distribution that approximates in form the topography of the mountain or mountainous area in which the population occurs (Fig. 1). Thus, the highest \( q_k \) values are exhibited by the Qomolungma population, a very flat \( q_k \) distribution is found in the Bolivian altiplano, a gentle double-breasted \( q_k \) distribution is shown by the unicorns that clamber about the Grand Tetons, and so on. Two possible explanations come to mind.

First, the congruence between frequency distribution and topography simply may reflect the generally harmonious way in which all unicorns interact with the rest of animate and inanimate nature (Johnsgard and Johnsgard 1982). Second, the montane unicorn may be trying to give us a gentle lesson. By distributing itself in concert with the generally sparse food supply in montane regions and as a way of minimizing the unicorns’ risk of puncturing each other with their horns. Horns are much longer in the montane unicorn (\( \mu = 2.3 \text{ m}, \sigma^2 = 0.3 \)) than in other species, as they are used to dig through the snow pack to obtain food in winter as well as for defense.

Cold temperatures and the presence of large predators tend to counter the influence of territoriality. Thus in the coldest region, on Qomolungma, we find the greatest degree of aggregation as indicated by the extended right tail of the \( q_k \) distribution (Fig. 1A) and the high \( I_{Mr} \) curve (Fig. 2A). Despite their intrinsic territoriality, Qomolungma unicorns spend much time huddled together in order to minimize heat loss. On the other hand for populations in the warmest areas, Machu Picchu and Torres del Paine, we find \( q_k \) distributions strongly truncated at \( k = 15 \) (Figs 1C, F) and with very low-lying \( I_{Mr} \) curves (Figs 2C, F). The two other populations are intermediate in all respects.

Only in two of the six regions are large predators a significant force. In the Qomolungma region the snow leopard (Panthera uncia) preys heavily on unicorns, and in the Torres del Paine region the puma (Felis concolor) does likewise. This may explain why unicorns were never observed in groups of less than six in these regions (Figs 1A, F). Others have observed that four adult unicorns usually can form a defensive circle capable of deflecting any attack by a lone snow leopard (della Roba 1985).

**Conclusions**

The analysis of dispersion patterns requires attention to three aspects of methodology: definition, selection of scale, and calculation of magnitude or degree. Use of such analyses to make inference about the characteristics (dimensions, spacing, etc.) of the actual aggregations or “patches” of organisms may also be a possibility. In this article I have discussed only certain of these methodological issues. Let me now put my recommendations in the context of the more general framework.

**Perspective on definitions**

I have argued the importance of developing a conceptual definition of aggregation and one that leads directly to an index or indices with a simple biological interpretation, e.g. one expressed in units relating directly to probability of co-occurrence or encounter, physical distance, etc. In principle the definition selected then determines both the index and the sampling or censusing procedures that should be used. In practice there are constraints that limit our choice of definitions. In the
ideal or “fully sampled” case we have a map showing the exact location of every individual organism in the population. This gives us the option of using any of a variety of methodologies (e.g. spectral analysis, pattern analysis, distance measurements, quadrat sampling) and hence equally unlimited freedom in selecting a definition. But rarely is this ideal case realizable except for areas that are very small or contain very few individuals. Likewise complete censusing of a grid of quadrats, though it proved feasible for unicorns on their various 10,000 km$^2$ ranges, will rarely be possible for other organisms except for very small grids. For highly mobile organisms, the instantaneous determination of position for even a sample of individuals usually will be difficult or impossible. Even for sessile organisms, measurement of distances to the $n$ nearest neighbors (from a randomly selected point or individual) is very difficult and time consuming in most field situations for high values of $n$ (e.g. $>3$). In sum, there are many operational constraints that can limit our approaches to the study of spatial pattern and to the defining and measuring of aggregation. There will be a continuing need for methods for analyzing “sparsely sampled patterns”.

The other definitional issue addressed here is whether the definition always need be one that results in all the information on dispersion patterns be boiled down to a single number, such as an $I_M$ value. A multidimensional definition requiring less severe condensation of the information, such as a plot of $I_M$ vs $r$ for each of a series of quadrat sizes (see below), often will be more useful.

Similar decisions are required when distance measurements are used rather than quadrat data. Though to my knowledge none have been proposed, it is possible to devise a family of indices that takes into account not only the mean distance to the $n$th nearest neighbor but also the frequency distribution of individual values about that mean. Such a family would express the same sort of information as does the $I_M$ family of indices calculated for a given quadrat size.

### Scales

Degree of aggregation in nature is always strongly a function of spatial scale. Any study that pretends to analyze the spatial patterns of a population must explicitly consider a wide range of spatial scales, if it is to avoid being simply another mathematical ecological exercise in ethereal scholasticism and irrelevance. By “consider” I mean that the study must at least discuss the distributional patterns present at each of a variety of scales. Thus my analysis of the unicorn populations based on use of a single quadrant size would be reprehensible if I were not using these data primarily to illustrate certain specific methodological problems.

Several approaches to the study of the variation of aggregation with scale have been devised (see Pielou 1977, Greig-Smith 1983, and references cited in Introduction). For studies using quadrat sampling, one can successively employ a range of quadrant sizes and for each size calculate $I_M$ or the $I_M$ curve. For studies using distance sampling, one can 1) successively calculate the distance between randomly selected points or organisms to their first, second, third, etc. nearest neighbors; 2) calculate the mean distances to first, second, etc. neighbors; and 3) divide each mean by the mean distance predicted on the assumption of a completely random distribution. If this ratio ($R$) is plotted versus neighbor rank, the resultant curve will contain essentially the same information as the curve obtained by plotting $I_M$ versus quadrat size. Either curve will be a good representation of the effect of scale on degree of aggregation. Neither approach is intrinsically superior despite occasional suggestions to the contrary (e.g. Paloheimo and Vukov 1976). Nevertheless it doubtless would be profitable to compare the results of the two approaches applied to the same sets of data. This has never been done.

### Components of patterns

Perhaps because there are two obvious steps to the quantification of aggregation – 1) selection of one or more spatial scales at which to make measurements and 2) calculation or graphical representation of the intensity or magnitude of aggregation, a few authors have been led to discuss the spatial patterns as if they themselves consisted of two fundamental components and to make unwarranted statements about exactly what particular indices do and do not measure.

Pielou (1977: 155) states this two-component idea most clearly: “A spatial pattern in a continuum obviously has two quite distinct aspects: they may be called intensity and grain”. She defines grain as the degree to which high density patches “are large in area and widely spaced” and intensity as “the extent to which density varies from place to place”. Then she claims indices such as $I_M$ “measure only the intensity, and not the grain, of a pattern”, which seems patently untrue. It would be highly unlikely, though possible in specific circumstances, to obtain the same $I_M$ values for two spatial patterns of equal “intensity”, one consisting of small patches closely spaced and the other consisting of large patches widely spaced. In other words, $I_M$ values are sensitive to such differences in “grain”. All aspects of pattern would seem to have the potential to influence values of $I_M$ regardless of the size of sampling unit employed, despite the claim (Pielou 1977: 156) that “indices of aggregation calculated from data obtained by sampling with quadrats of one size are all measures of the intensity of a pattern”.

Is it possible to dissect spatial patterns into a finite number of components? In principle, it may be possible to come up with an operational definition of a “patch” that allows any spatial distribution to be viewed as a number of the patches. This suggests the possibility of describing the population’s spatial distribution in terms of patch characteristics such as: size and shape of area...
occupied, intrapatch density, sharpness of patch boundaries, distance between patches, and so on (e.g. Goodall and West 1979). The basic list clearly contains more than two "distinct aspects", however; the grain-intensity dichotomy is an arbitrary and artificial one. Furthermore, two additional aspects of pattern are superimposed on these basic ones. First, for each of the above patch characteristics one must consider not only its mean value but also its variability. For example, degree of aggregation will be a function not just of mean patch size but also of the frequency distribution of patch sizes. Secondly, a hierarchy of patches may exist, with small patches grouped in various ways to form larger patches that form still larger patches and so on. At each level in such a hierarchy, one can define all the same patch characteristics listed above and determine the mean, variance, etc. for each.

Given this reality, it is unrealistic to expect that mathematical approaches can be developed that are capable of sorting out and quantifying the different properties and scales of patches in natural populations. For that purpose, observant eyes and good graphs and maps will remain the indispensable tools. On the other hand any quantifiable property of a spatial distribution always can be measured using a wide range of sampling scales. If the property is biologically meaningful or interpretable, then examination of the pattern of its variation with sampling scale should be useful.

Acknowledgements - Thanks are expressed to D. Bartley and J. Varnell for help in finding unicorian distributions and calculating their properties and to B. D. Collier, E. D. Ford, D. J. Varnell for help in finding unicorian distributions and calculating their properties. This paper is dedicated to the memory of the late L. C. Cole and F. V. della Roba, premier monocerologists of the 20th century.

References


Shepard, O. 1967. The lore of the unicorn. – Barnes and Noble, New York.


