

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/7756538>

# Properties of crowding indices and statistical tools to analyze parasite crowding data

Article in *Journal of Parasitology* · May 2005

DOI: 10.1645/GE-281R1 · Source: PubMed

CITATIONS

31

READS

66

4 authors:



Jenő Reiczigel

University of Veterinary Medicine Budapest

106 PUBLICATIONS 2,638 CITATIONS

SEE PROFILE



Zsolt Lang

University of Veterinary Medicine Budapest

46 PUBLICATIONS 378 CITATIONS

SEE PROFILE



Lajos Rozsa

MTA, ÖK

77 PUBLICATIONS 2,121 CITATIONS

SEE PROFILE



Béla Tóthmérés

University of Debrecen

321 PUBLICATIONS 5,895 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Pentastomiasis [View project](#)



Human Impacts on Terrestrial Biodiversity [View project](#)

## PROPERTIES OF CROWDING INDICES AND STATISTICAL TOOLS TO ANALYZE PARASITE CROWDING DATA

Jenő Reiczigel, Zsolt Lang\*, Lajos Rózsa†, and Béla Tóthmérész‡

Department of Biomathematics and Informatics, Faculty of Veterinary Science, Szent István University, P.O. Box 2, Budapest H-1400, Hungary.  
 e-mail: Reiczigel.Jeno@aotk.szie.hu

**ABSTRACT:** Crowding, i.e., the size of the infrapopulation inhabiting an individual host, is a major component of parasites' environment, which often influences both morphological and life-history characters (the so-called density-dependent characters) in different parasite taxa. Although crowding equals intensity in case of a single parasite individual, mean intensity of the host population does not define mean crowding of the parasite population. Crowding indices are notoriously hard to handle statistically because of the inherently large number of nonindependent values in data. In this study, we aim to investigate the apparently paradox features of crowding indices and to make some proposals and also to introduce statistical methods to calculate confidence intervals and 1-sample and 2-sample tests for mean crowding. All methods described in this study are supported by the freely distributed statistical software Quantitative Parasitology.

Traditionally, parasitologists used to claim that the host's body is the single most important component of parasite's environment. However, since Read's (1951) study on the "crowding effect," there is a slowly increasing interest in changing this view and in investigating potential effects of infrapopulation size on the ecology, evolution, and behavior of parasites. Infrapopulation size often correlates negatively with helminth body size both in the intermediate (Keymer, 1981) and in the definitive hosts (see Roberts, 1961; Shostak and Dick, 1987). Density-dependent intraspecific variability of helminth body size may be dramatic because the largest individuals may be 90 times the size of the smallest (Szalay and Dick, 1989). Similar effects on helminth body size and fecundity also occur in multispecies helminth assemblages (Dezfuli et al., 2002). Body size differences often mirror differences in life-history traits and particularly in fecundity; thus, helminth fecundity is often density dependent also (Zervos, 1988a, 1988b). Infrapopulation size may also correlate with measures of sexual selection in parasites, such as sex ratio in protists (Read et al., 1992, 1995), nematodes (Poulin, 1997), spiny-headed worms (Poulin and Morand, 2000; Sasal et al., 2000), and even arthropods (Rózsa et al., 1996; Rózsa, 1997).

Therefore, future students of parasite ecology, evolution, and behavior are likely to face the challenge of quantifying measures of parasite crowding and comparing these measures statistically across different samples. Crowding is defined as intensity, or infrapopulation size, from the parasite's point of view. This means that crowding equals intensity when speaking about an individual parasite or an infrapopulation living on a single host. In contrast, mean intensity in a host population does not predict mean crowding in the corresponding parasite population. Mathematically, mean host intensity is the sum of intensity values over the hosts divided by the number of hosts, whereas mean crowding is the sum of the crowding values over the parasites divided by the number of parasites. Let us have 2 samples consisting of 3 hosts each to illustrate this point:

	intensities	mean (median) intensity	mean (median) crowding
Sample A:	4, 5, 6	$\frac{4 + 5 + 6}{3} = 5$ (5)	$\frac{4 \times 4 + 5 \times 5 + 6 \times 6}{4 + 5 + 6} = 5.13$ (5)
Sample B:	1, 4, 7	$\frac{1 + 4 + 7}{3} = 4$ (4)	$\frac{1 \times 1 + 4 \times 4 + 7 \times 7}{1 + 4 + 7} = 5.5$ (7)

In this hypothetical example, hosts tend to harbor larger infrapopulations in sample A, whereas parasites tend to live in larger infrapopulations in sample B. This apparent contradiction is a consequence of a difference in the aggregation of samples. Interestingly, average individual parasites come from hosts with intensities higher than the average intensity, and this shift is more pronounced in more aggregated samples.

Empirical data (Fig. 1) suggest that mean intensity may roughly predict mean crowding. However, this pattern is valid only within a narrow taxonomic unit, and contradictory situations, i.e., one species having larger mean intensity and the other having larger mean crowding, may still occur. Thus, when studying the influence of crowding on the ecology of parasites, statistical analysis should be performed directly on measures of crowding.

Similar approaches have been proposed in spatial and behavioral ecology (Lloyd, 1967; Jarman, 1974). Spatial ecologists use artificial sampling units (so-called quadrats) to determine abundance or density of species; thus, their sampling methods focus on units potentially harboring several individuals belonging to different species. The situation is similar in parasitology, with the major difference being that host individuals act as natural sampling units in this case. For use in spatial ecology, Lloyd (1967) introduced 2 types of crowding indices, namely 'mean crowding' and 'mean demand,' to derive an index of aggregation, called 'index of patchiness.' Lloyd's mean crowding is the mean number per individual of other individuals in the same quadrat:

$$m^* = \frac{\sum_{j=1}^N (X_j - 1)}{N}$$

where  $N$  is the number of all individuals and  $X_j$  ( $j = 1, 2, \dots, N$ ) is the number of conspecific individuals living in the same quadrat as the  $j$ th individual. Apparently, this idea relies on the assumption that an individual does not 'crowd' itself, suggest-

Received 17 November 2003; revised 14 June 2004; 27 July 2004; accepted 16 August 2004.

\* Nomogram Ltd., Arany János utca 70, Budapest H-1165, Hungary.

† Animal Ecology Research Group of the Hungarian Academy of Sciences, Hungarian Natural History Museum, Ludovika tér 2, Budapest H-1083, Hungary.

‡ Department of Ecology, Faculty of Sciences, University of Debrecen, Debrecen H-4010, Hungary.

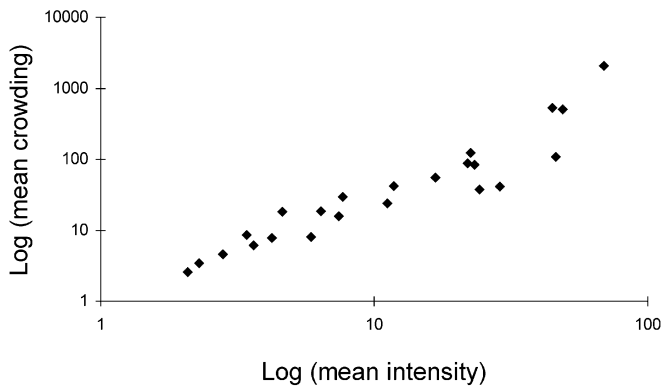


FIGURE 1. Mean intensity correlates positively with mean crowding in case of empirical data of avian lice (data were obtained from Rózsa et al., 2000, and references therein).

ing that ‘crowding effects’ are realized through interactions among individuals, such as aggression. On the other hand, mean demand is based on the presumption that individuals do crowd themselves even when they are alone, provided that crowding effects are realized through depletion of limited resources. Thus, mean demand equals the mean number of individuals per quadrat per individual. By definition, Lloyd’s index of mean crowding equals 0, whereas mean demand equals + 1 in a lonely individual, and mean demand equals mean crowding + 1 in general. Note that mean demand is identical with the index of ‘typical group size’ as defined by Jarman (1974) to quantify the group size within which an average individual lives.

Any of these indices could be used to analyze density-dependent effects in parasites. However, one might choose to be interested in analyzing crowding effects considering other scales, e.g., the logarithmic scale, expressing the biological notion that proportional changes are more biologically meaningful than additive ones. In quantification and crowding indices we define parasite crowding in general as a monotone increasing function of the infrapopulation size (intensity), which incorporates the above crowding indices as special cases.

Given a wide range of modern statistical methods for quantifying parasites in samples of hosts (Rózsa et al., 2000), it is surprising to note that none of the standard methods can be used to handle measures of crowding. All available statistical tools used to compare parasite burdens across samples are based on the presumption that data consist of statistically independent events, whereas crowding data almost never meet this assumption. To illustrate our point, consider the following hypothetical sample consisting of 3 hosts:

host intensity values: 1, 2, 3 (mean intensity = 2)  
 parasite crowding values: 1, 2, 2, 3, 3, 3  
 (mean crowding = 2.33)

Let us eliminate the last parasite individual to perceive how the data change:

host intensity values: 1, 2, 2 (mean intensity = 1.66)  
 parasite crowding values: 1, 2, 2, 2, 2  
 (mean crowding = 1.8)

Alternatively, let us eliminate the last host individual to perceive how the data change:

host intensity values: 1, 2 (mean intensity = 1.5)  
 parasite crowding values: 1, 2, 2 (mean crowding = 1.66)

When a single parasite or host is added or eliminated, there is only a single change in the intensity data, but there are several parallel changes in crowding data. This means that crowding data consist of nonindependent values, which exhibit multiple changes synchronously because of a single event. Statistical methods currently used in parasitology are incapable of handling this effect. Of course, this problem does not emerge in samples where individual intensities do not exceed 1, such as in many host–parasitoid systems. However, intensity may exceed 1 in the vast majority of host–parasite systems; thus, the nonindependence of crowding data is a real problem in parasite ecology. The aim of this study is to investigate some curious features of crowding data and to introduce statistical methods that are capable of handling their nonindependence. Mathematical definition is given in section 2, some interesting properties are described in section 3, statistical inference is outlined in section 4 followed by discussion and concluding remarks in section 5.

### QUANTIFICATION AND CROWDING INDICES

For an individual parasite, crowding can be defined simply by intensity, i.e., the size of the infrapopulation it lives in. In some circumstances, rescaling this by a monotone function may offer a more meaningful interpretation; this also depends on the type of mathematical model in which crowding is to be analyzed. An example of this is the logarithmic transformation, i.e., taking crowding = ln(intensity), which expresses the view that relative differences are more appropriate to consider than absolute ones; thus, the difference between 1 and 2 is the same as between 10 and 20. Another example of such a transformation is Lloyd’s definition of mean crowding, where crowding = intensity – 1, resulting in 0 crowding for individuals living alone.

To quantify the level of crowding in a parasite population or sample, one can use a statistic (a so-called ‘‘measure of location’’) calculated from the individual values. The most widely used measures of location are the mode, the median, and the mean. (The fourth potential measure of location, the geometric mean, is equivalent to using the logarithmic scale and applying the mean.) However, in case of highly skewed intensity distributions, which are typical of many parasites, there is a good chance that the mode or median of crowding is equal to the largest value. As an example, assume a sample of 12 hosts with intensities 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 4, 20 (mean intensity = 3.25), which corresponds to a sample of 39 parasites, with mean = 11.36, and mode = median = 20. The bar charts in Figure 2 illustrate the difference between intensity and crowding.

The main problem is that in this case both the mode and the median rely on data from a single host, which may dramatically influence their sampling properties such as SE. In addition, these measures are uncomfortable to handle statistically when having many tied values, so we propose to use the mean. Mean crowding in a population or sample of *P* parasites is defined as

$$mean\ C = \frac{\sum_{j=1}^P c_j}{P},$$

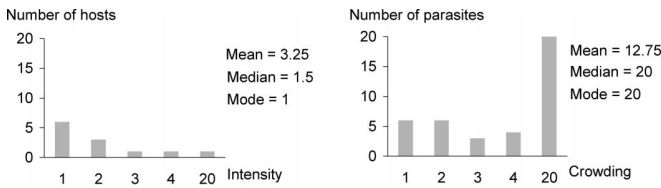


FIGURE 2. Frequency distribution of intensity and crowding for the same hypothetical sample. Intensity is 20 for a single host, but this results in crowding = 20 for 20 parasites.

where  $c_j$  denotes crowding value of the  $j$ th parasite. To make a clear distinction between sample and population mean, we will follow the convention of denoting the population mean by  $E(C)$  and the sample mean by  $\bar{c}$ . This is in accordance with the usual statistical notation and helps avoid confusion when the sample mean  $\bar{c}$  is to be used as an estimate of the unknown population mean  $E(C)$ .

It is worth noting how mean crowding is related to indices of intensity. Assume that the host population consists of  $H$  hosts, and let  $i_h$  denote the intensity for the  $h$ th host ( $h = 1, 2, \dots, H$ ). Then the parasite population consists of  $P = \sum_{h=1}^H i_h$  individuals, and the population mean crowding can be written as

$$E(C) = \frac{\sum_{h=1}^H i_h^2}{\sum_{h=1}^H i_h}$$

given crowding is quantified directly by the size of the parasite infrapopulation (Lloyd, 1967; Jarman, 1974). If using a general monotone increasing function of intensity  $f(i)$  to measure crowding, the formula changes to

$$E(C_f) = \frac{\sum_{h=1}^H i_h f(i_h)}{\sum_{h=1}^H i_h}$$

Especially for the logarithmic scale, it results in

$$E(C_{\log}) = \frac{\sum_{h=1}^H i_h \ln(i_h)}{\sum_{h=1}^H i_h}$$

Mean crowding can be defined in a similar manner for theoretical probability distributions, sometimes called “infinite populations.” Especially, for a parametric distribution, mean crowding can be expressed as a function of the parameters. In case of a negative binomial, which is the most widely used distributional model ever since Crofton (1971), this results in

$$E(C) = m + m/k + 1 \tag{1}$$

where  $m$  and  $k$  denote the mean value and the negative binomial exponent, respectively (Lloyd, 1967).

**PROPERTIES OF MEAN CROWDING**

Mean crowding in a population is always greater than or equal to mean intensity. Equality holds only if all hosts have the same number of parasites, which is nonrealistic for real

parasites. The greater the variance of intensity, the greater the difference between mean crowding and mean intensity. Actually, the following relation holds

$$E(C) = E(I) + \frac{\text{var}(I)}{E(I)} \tag{2}$$

where  $E(I)$  and  $\text{var}(I)$  denote the population mean and variance of intensity, respectively (Lloyd, 1967). This result implies that the mean and variance of intensity fully predict mean crowding. Unfortunately, log-scaled mean crowding cannot be calculated directly from the mean and variance of intensity.

If mean crowding is known in each one of  $r$  populations, then crowding in the pooled population is equal to the weighted average of the individual mean crowdings with weights being the numbers of parasites in the corresponding populations. That is, given the crowding indices and the number of parasites  $C_m$  and  $P_m$  ( $m = 1, 2, \dots, r$ ).

$$E(C)_{\text{pooled}} = \frac{\sum_{m=1}^r P_m C_m}{\sum_{m=1}^r P_m}$$

From this, it follows that the pooled mean crowding is between the minimal and maximal one, and especially, if all crowdings are equal, then the pooled one is also equal to this common value.

There are some surprising properties of mean crowding, which may seem to be paradoxical at first sight, e.g., increasing the number of parasites in particular hosts may decrease the population mean of crowding, or vice versa. To illustrate this, assume that the host population consists of 3 hosts as follows:

intensity: 1, 2, 10:

$$E(I) = 4.33,$$

$$E(C) = \frac{1 \times 1 + 2 \times 2 + 10 \times 10}{1 + 2 + 10} = 8.077.$$

Adding a parasite to the second infrapopulation, we get

intensity: 1, 3, 10:

$$E(I) = 4.66,$$

$$E(C) = \frac{1 \times 1 + 3 \times 3 + 10 \times 10}{1 + 3 + 10} = 7.857,$$

i.e., the mean crowding decreases. The resolution of this paradox is that we have introduced the new individual into an infrapopulation in which crowding was well below the population mean. Although we have increased crowding here, it still remained below the mean, so we actually increased the number of individuals with a crowding value below the population mean. Formula (2) above offers another argument, namely, that the modified host population has higher mean intensity but less variance of intensity. Thus, the first term of the formula has increased, whereas the second term has decreased, resulting in an overall decrease of mean crowding.

Another apparently paradox property highlights the difference between crowding and intensity. Adding or removing a

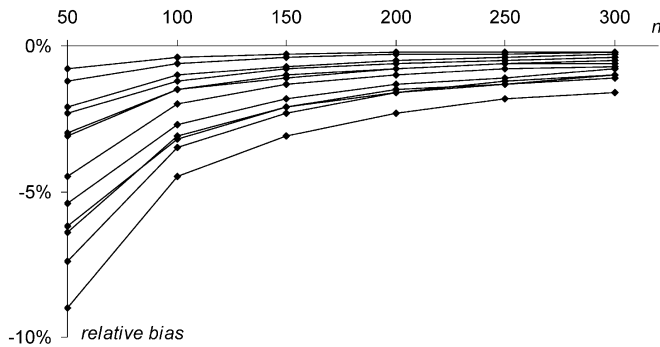


FIGURE 3. Dependence of the relative bias (bias divided by the true value) of mean crowding on sample size, estimated for each of the 12 empirical distributions (listed in Table I) and for each sample size from 10,000 simulated samples. Asymptotically, relative bias is proportional to  $1/n$  where  $n$  is the sample size.

host and its parasites may increase mean crowding and decrease mean intensity (or vice versa) at the same time. An example:

intensity: 1, 10:

$$E(I) = 5.5,$$

$$E(C) = \frac{1 \times 1 + 10 \times 10}{1 + 10} = 9.182,$$

intensity: 1, 7, 10:

$$E(I) = 6,$$

$$E(C) = \frac{1 \times 1 + 7 \times 7 + 10 \times 10}{1 + 7 + 10} = 8.333,$$

The explanation is simple. Looking at the first population, we can see that its mean crowding (9.182) is higher than its mean intensity (5.5). Introducing a host with intensity 7, which is between mean intensity and mean crowding, mean intensity will increase whereas mean crowding will decrease.

### STATISTICAL INFERENCE

In the following, estimation and testing of mean crowding are addressed. A point estimate and a confidence interval is introduced for the 1-sample problem. Concerning statistical testing, both 1- and 2-sample tests are derived from the 1-sample confidence interval.

The main difficulty is that the usual sampling method is sampling the hosts because one cannot draw a random sample from the parasite population without catching the hosts they live in; furthermore, when a host is sampled, all of its parasites must be sampled to determine the infrapopulation size. Thus, sampling the hosts necessarily results in a cluster sample of parasites, making statistical inference a little more difficult than it would be if a random sample could be drawn directly from the parasite population.

*Point estimate:* The simplest method to estimate mean crowding is to apply the same formula (2) for the sample that was

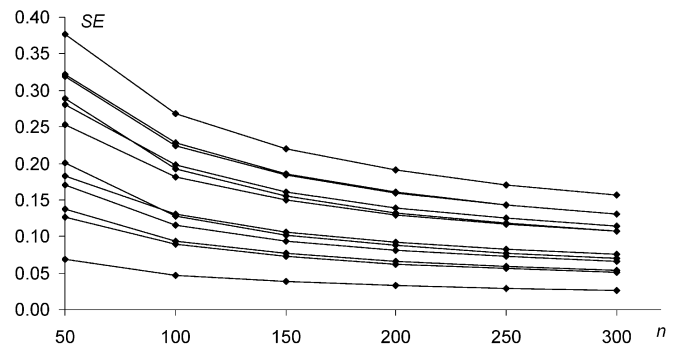


FIGURE 4. Dependence of SE of mean crowding on sample size. SE is estimated for each of the 12 empirical distributions and for each sample size from 10,000 simulated samples.

used for the population (this kind of estimate is called a plug-in estimate), which results in

$$\bar{c} = \frac{\sum_{j=1}^n i_j^2}{\sum_{j=1}^n i_j},$$

where  $n$  is the number of hosts in the sample, and  $i_j$  denotes the intensity for the  $k$ th host ( $j = 1, 2, \dots, n$ ). The sample mean crowding  $\bar{c}$  has a downward bias, i.e., it tendentially underestimates the population mean crowding  $E(C)$ . (For the mathematical proof see the Appendix. Note that for the negative binomial distribution, it was proven by Lloyd, 1967). The bias is a consequence of sampling the hosts. The sample mean based on a random sample from the parasite population would be unbiased. Fortunately, as being a so-called ratio estimate,  $\bar{c}$  is asymptotically unbiased, i.e., its bias tends to 0 (of the order of  $1/n$ ) if the sample size tends to infinity (Cochran, 1963, p. 161). However, it is not clear whether the bias is relevant for a particular sample size. It is not easy to make a general statement about this because it depends on the distribution of intensity. Therefore, we investigate this by simulation based on various realistic intensity distributions (Fig. 3).

The estimate  $\bar{c}$  has further nice asymptotic properties. It is asymptotically normally distributed and its SE tends to 0, with a magnitude of  $1/\sqrt{n}$ , if the sample size tends to infinity (Fig. 4), see Cochran (1963, pp. 157–158).

Assuming a parametric distribution for intensity, it is possible to construct a maximum likelihood (ML) estimate for mean crowding, which is expected to be better than  $\bar{c}$ . The usual parametric model for parasite intensity distributions is the negative binomial, although in some cases this model does not fit either (Rékási et al., 1997). The ML estimate for mean crowding is obtained according to (formula 1) from the ML estimates of mean intensity and the negative binomial exponent. The latter can be calculated following Bliss and Fisher (1953). For the negative binomial distribution, all these results as well as the calculation of confidence intervals based on them is described by Lloyd (1967). Note that (formula 2) can also be applied to estimate mean crowding from the sample mean and variance of intensity, enabling estimation of mean crowding from published data, as mean and SD of intensity are given in most studies.

TABLE I. Coverage probabilities of the BCa confidence interval for different distributions and sample sizes, estimated from 10,000 simulated samples. Coverage values too low compared with the nominal level ( $\leq 0.88$ ,  $\leq 0.94$ ,  $\leq 0.97$  for nominal .90, .95, .975, respectively) are typed in boldface.

Parasite species	True mean		n = 100			n = 200			n = 300		
	Intensity	Crowding	90%	95%	97.5%	90%	95%	97.5%	90%	95%	97.5%
<i>Anatoecus dentatus</i>	17.0	51.6	0.904	0.951	0.976	0.900	0.948	0.975	0.898	0.949	0.975
<i>A. icterodes</i>	62.2	491	0.883	<b>0.928</b>	<b>0.952</b>	0.899	0.948	0.972	0.907	0.957	0.979
<i>Brueelia tasniamae</i>	44.2	108	0.907	0.956	0.979	0.908	0.953	0.978	0.897	0.947	0.975
<i>Carnus hemapterus</i>	29.1	53.1	0.899	0.945	<b>0.966</b>	0.909	0.954	0.974	0.906	0.954	0.976
<i>Coloceras softoticus</i>	5.23	19.4	0.890	<b>0.930</b>	<b>0.950</b>	0.898	0.950	0.971	0.911	0.956	0.978
<i>Columbicola bacillus</i>	7.64	16.6	<b>0.855</b>	<b>0.900</b>	<b>0.922</b>	0.892	<b>0.935</b>	<b>0.959</b>	0.902	0.949	0.971
<i>Cummingsiella aurea</i>	24.7	138	<b>0.842</b>	<b>0.883</b>	<b>0.899</b>	0.892	<b>0.938</b>	<b>0.958</b>	0.901	0.948	<b>0.968</b>
<i>Myrsidea cornicis</i>	12.5	45.4	<b>0.877</b>	<b>0.921</b>	<b>0.942</b>	0.902	0.950	0.973	0.899	0.950	0.973
<i>M. isostoma</i>	23.7	97.5	<b>0.864</b>	<b>0.908</b>	<b>0.929</b>	0.897	0.944	<b>0.967</b>	0.902	0.952	0.972
<i>Philopterus atratus</i>	24.0	85.0	0.904	0.950	0.972	0.900	0.951	0.975	0.906	0.957	0.980
<i>P. ocellatus</i>	11.5	25.7	<b>0.868</b>	<b>0.916</b>	<b>0.944</b>	0.897	0.944	<b>0.966</b>	0.904	0.948	0.971
<i>Rhynonirmus helvolus</i>	62.8	505	<b>0.865</b>	<b>0.905</b>	<b>0.925</b>	0.900	0.945	<b>0.967</b>	0.896	0.951	0.974
Neg. bin. $m = 2, k = 0.2$	5.16	11.8	<b>0.876</b>	<b>0.918</b>	<b>0.944</b>	0.897	0.946	<b>0.969</b>	0.910	0.954	0.976
Neg. bin. $m = 2, k = 0.1$	7.45	20.4	<b>0.868</b>	<b>0.909</b>	<b>0.935</b>	0.902	0.948	<b>0.968</b>	0.905	0.951	0.976
Neg. bin. $m = 2, k = 0.05$	11.5	37.0	<b>0.857</b>	<b>0.903</b>	<b>0.929</b>	0.899	0.946	<b>0.967</b>	0.908	0.952	0.977
Neg. bin. $m = 20, k = 0.2$	32.4	106	<b>0.863</b>	<b>0.907</b>	<b>0.931</b>	0.898	0.945	<b>0.967</b>	0.912	0.954	0.977
Neg. bin. $m = 20, k = 0.1$	47.0	186	<b>0.853</b>	<b>0.896</b>	<b>0.921</b>	0.897	0.943	<b>0.965</b>	0.909	0.952	0.978
Neg. bin. $m = 20, k = 0.05$	74.1	343	<b>0.847</b>	<b>0.896</b>	<b>0.921</b>	0.906	0.944	<b>0.964</b>	0.905	0.950	0.974

Neg. bin. = Negative binomial distributions.

*Confidence interval construction:* Because of the heterogeneity and typical nonnormality of parasite distributions, confidence interval construction requires application of nonparametric methods or bootstrap techniques. As was mentioned above, sample mean crowding is a biased estimate of the population mean. Furthermore, sample mean and sample variance of crowding are correlated. In such circumstances, the bias-corrected and accelerated (BCa) bootstrap confidence interval appears to be the most appropriate choice for confidence interval construction (Efron and Tibshirani, 1993). Although there are some asymptotic results concerning this method (Davison and Hinkley, 1997), investigating its performance under realistic circumstances is crucial. Because of the heterogeneity of the distributions, it is difficult to use theoretical methods for this purpose; therefore, we preferred to carry out a simulation study based on typical parasite distributions and realistic sample sizes.

For this purpose, we chose 18 distributions, including 12 empirical distributions of avian ectoparasites (for the sources of data, see Rózsa et al. [2000] and references therein) and 6 different negative binomials. Distributions derived from empirical data were represented by 300 intensity values corresponding to the equidistant quantiles (1/300 quantiles) of the smoothed empirical distribution. Negative binomial distributions were truncated at their upper 0.99 quantiles because occurrence of unrealistically high intensity values attributed to unboundedness of the negative binomial distribution would influence heavily the estimate of crowding. We used these distributions as “populations,” calculated their crowding values, then drew samples from each of them and constructed BCa confidence intervals from the samples. Because the population crowding was known, it was possible to determine the actual coverage probabilities of the confidence intervals and to compare the actual coverage to the nominal one. We did this with nominal 90, 95, and 97.5% intervals (97.5% will be needed for a 95% 2-sample

test, see below). To provide sufficient precision, coverage was estimated from 10,000 simulated samples for each distribution, resulting in SE of the coverage estimate  $\leq 0.3, 0.22, 0.16\%$  at level of 90, 95, 97.5%, respectively.

We found that for sample sizes below 100, the actual coverage reaches the nominal one only for the least aggregated distributions. For typically distributed ectoparasites sample sizes of 100–200 are needed, whereas for highly aggregated ones 300 or more. Evidently, if 2–3% of the hosts harbor about 50% of parasites, a small sample is unlikely to contain any of these host individuals; thus, any estimate will be extremely biased. On the other hand, the BCa confidence interval seems to be correct for large samples (Table I). A similar simulation study was carried using the same distributions, but calculating log-transformed crowding values. In this case, smaller sample sizes were found to be sufficient (100 for less aggregated and about 200 for more aggregated distributions).

*Statistical testing:* An 1-sample test for mean crowding can be derived from the above confidence interval in the usual manner. The null hypothesis stating that mean crowding in the study population equals to a certain hypothetical value  $C_0$ , i.e.,  $H_0: E(C) = C_0$  against the alternative of inequality, is to be rejected at level  $\alpha$  if the  $(1 - \alpha)$ -level confidence interval for  $E(C)$  does not contain the hypothetical value  $C_0$ .

Similarly, an  $\alpha$  level-2-sample test can be based on 2  $(1 - \alpha/2)$ -level confidence intervals. The null hypothesis that crowding is equal in 2 populations, i.e.,  $H_0: E(C_1) = E(C_2)$  against the alternative of inequality, is to be rejected at level  $\alpha$  if the  $(1 - \alpha/2)$ -level confidence intervals for  $E(C_1)$  and  $E(C_2)$  do not overlap. Especially, 2 samples with nonoverlapping 97.5% confidence intervals for mean crowding provide evidence for the difference between population mean crowdings at  $P = 0.05$ .

One-tailed tests, i.e., those with directional alternative hy-

TABLE II. Coverage probabilities of the jackknife confidence interval by Reed (1983), estimated from 10,000 simulated samples. Coverage values too low compared with the nominal level ( $\leq 0.88$ ,  $\leq 0.94$ ,  $\leq 0.97$  for nominal .90, .95, .975, respectively) are typed in boldface.

Parasite species	n = 200			n = 300		
	90%	95%	97.5%	90%	95%	97.5%
<i>Anatoecus dentatus</i>	0.907	0.957	0.978	0.907	0.956	0.978
<i>A. icterodes</i>	0.898	0.946	0.972	0.894	0.944	0.973
<i>Brueelia tasmiana</i>	0.900	0.950	0.972	0.900	0.949	0.973
<i>Carnus hemapterus</i>	0.892	0.942	<b>0.967</b>	0.902	0.948	0.972
<i>Coloceras sofiticus</i>	0.895	0.941	<b>0.966</b>	0.903	0.948	0.970
<i>Columbicola bacillus</i>	<b>0.869</b>	<b>0.914</b>	<b>0.936</b>	0.882	<b>0.926</b>	<b>0.950</b>
<i>Cummingsiella aurea</i>	<b>0.874</b>	<b>0.917</b>	<b>0.942</b>	0.892	<b>0.936</b>	<b>0.958</b>
<i>Myrsidea cornicis</i>	0.884	<b>0.933</b>	<b>0.958</b>	0.891	<b>0.939</b>	<b>0.963</b>
<i>M. isostoma</i>	0.885	<b>0.932</b>	<b>0.955</b>	0.898	0.941	<b>0.967</b>
<i>Philopterus atratus</i>	0.895	0.944	<b>0.969</b>	0.910	0.954	0.978
<i>P. ocellatus</i>	0.893	<b>0.937</b>	<b>0.962</b>	0.899	0.944	<b>0.968</b>
<i>Rhynonirmus helvolus</i>	0.890	<b>0.934</b>	<b>0.959</b>	0.889	0.941	<b>0.966</b>
Neg. bin. $m = 2, k = 0.2$	0.892	<b>0.936</b>	<b>0.959</b>	0.898	<b>0.940</b>	<b>0.966</b>
Neg. bin. $m = 2, k = 0.1$	0.885	<b>0.933</b>	<b>0.957</b>	0.890	<b>0.938</b>	<b>0.960</b>
Neg. bin. $m = 2, k = 0.05$	<b>0.879</b>	<b>0.928</b>	<b>0.953</b>	0.899	<b>0.940</b>	<b>0.964</b>
Neg. bin. $m = 20, k = 0.2$	0.884	<b>0.933</b>	<b>0.955</b>	0.890	<b>0.936</b>	<b>0.961</b>
Neg. bin. $m = 20, k = 0.1$	0.890	<b>0.931</b>	<b>0.954</b>	0.896	<b>0.940</b>	<b>0.964</b>
Neg. bin. $m = 20, k = 0.05$	0.888	<b>0.932</b>	<b>0.953</b>	0.888	<b>0.936</b>	<b>0.961</b>

Neg. bin. = Negative binomial distributions.

potheses, such as, for example, "crowding in population 1 is greater than in population 2," can be constructed in the same manner but using 1-sided confidence intervals. A  $(1 - \alpha)$ -level 1-sided confidence interval can be obtained from a  $(1 - 2\alpha)$ -level 2-sided one by expanding it to infinity (or minus infinity) on 1 side, e.g., a 90% 2-sided interval  $(c_1, c_2)$  leads to the following 95% 1-sided intervals:  $(c_1, \infty)$  or  $(-\infty, c_2)$ .

The properties of the above tests follow from the properties of the corresponding confidence intervals. User-friendly statistical tools to estimate and to compare mean crowding are incorporated in the latest version Quantitative Parasitology, a software freely distributed through the authors' web page (Reiczigel and Rózsa, 2003).

## DISCUSSION

In recent decades, there was a sudden increase in the statistical tools used in spatial ecology. Although some of these methods could be used fruitfully in many other branches of sciences, the adoption of them is still ahead in many areas. Parasite ecologists in general have been failing to adopt Lloyd's indices, but behavioral ecologists have apparently also missed the point. Thus, Jarman repeatedly emphasized that group size is a major factor in an individuals' social environment and introduced an index to quantify typical group size (Jarman, 1974, 1982; P. J. Jarman and M. V. Jarman, 1979). By this term, he meant a group size typical to an 'average individual,' without realizing that this index was identical to Lloyd's index of mean demand. Despite the formal parallelism between these problems (sampling units: quadrat = group = infrapopulation inhabiting a host individual), it appears that nobody except Jarman (1974, 1982) proposed the use of crowding indices for the analysis of social behavior or parasite crowding effects yet.

Lloyd (1967) outlined parametric inference for the 1-sample problem, an estimation of SE, and construction of confidence

limits on the basis of this notion, assuming negative binomial distribution. For the general case, Reed (1983) proposed the use of jackknife to obtain confidence intervals for mean crowding. In this study, we applied the bootstrap, which is known to perform better in general than the jackknife. According to the simulation results shown in Table II, this holds in this specific case as well: bootstrap coverage with samples of 200 is better than jackknife coverage with samples of 300. Bootstrap intervals are superior not only with respect to coverage but also shorter by 1–4% than jackknife intervals. But our results suggest that even if applying the bootstrap method, samples of hundreds of hosts are necessary for a valid statistical inference in case of realistic parasite distributions. Sample size needed also depends on the scaling (linear, logarithmic, or other) of the crowding index.

The size of the parasite infrapopulation is a major determinant of the environment of a parasite individual. The potential physiological causes (competition for carbohydrates or the effect of growth inhibitors excreted by conspecifics or both) and morphological consequences of crowding had been intensively studied in cestodes for decades (see Roberts, 2000, for a recent review). A diverse set of further parasites, including protists, nematodes, acanthocephalans, and arthropods, also exhibit density-dependent variability in sex ratio, morphology, and life history. We hope that the presented statistical methodology will help refine the analysis of crowding data. Most previous authors who reported on the interaction between crowding and other parasite characters did in fact quantify intensity rather than crowding. Because mean intensity tends to predict mean crowding (Fig. 1), this approach is not at all fruitless. However, stronger correlation with and better prediction of parasite characters could be expected by introducing crowding into the model in place of intensity, as a consequence of averaging on parasites rather than on hosts.

To our best knowledge, this study is the first to introduce the

application of these indices in parasitology, and also to outline statistical tools to handle parasite-crowding data. We did our best to provide free access to user-friendly software that helps nonstatistician readers, and parasitologists in particular, to exhibit these calculations easily. We propose that not only parasite ecologists but parasite taxonomists also may benefit from the use of the above methods in the future. Because parasite morphology often co-varies with parasite crowding, it is advisable for the taxonomist to include crowding measures into parasite species descriptions. Obviously, if a species description is based on parasites collected from 1 or a few host individuals, this equals with providing the intensity data for each host. On the other hand, if parasites were collected from several hosts, mean crowding and its confidence interval should become an integral part of future species descriptions.

### ACKNOWLEDGMENTS

Helpful comments from 2 anonymous referees greatly contributed to this study. One of them made us understand that we were rediscovering Lloyd's indices published 33 yr before and even provided detailed advice on how to fit our current ideas about parasite crowding to the established methodology of spatial ecology. Peter J. Jarman kindly shared his ideas and reprints on typical group size with us. This study was supported by a grant from the Hungarian Scientific Research Fund (Grant T 035150).

### LITERATURE CITED

- BLISS, C. I., AND R. A. FISHER. 1953. Fitting the negative binomial distribution to biological data. *Biometrics* **9**: 176–200.
- COCHRAN, W. G. 1963. Sampling techniques, 2nd ed. Wiley and Sons, New York, 413 p.
- CROFTON, H. D. 1971. A quantitative approach to parasitism. *Parasitology* **62**: 179–193.
- DAVISON, A. C., AND D. V. HINKLEY. 1997. Bootstrap methods and their application. Cambridge University Press, Cambridge, U.K., 582 p.
- DEZFULI, B. S., S. VOLPONI, I. BELTRAMI, AND R. POULIN. 2002. Intra- and interspecific density-dependent effects on growth in helminth parasites of the cormorant, *Phalacrocorax carbo sinensis*. *Parasitology* **124**: 537–544.
- EFRON, B., AND R. J. TIBSHIRANI. 1993. An introduction to the bootstrap. Chapman and Hall/CRC Press, Boca Raton, Florida, 436 p.
- JARMAN, P. J. 1974. The social organisation of antelope in relation to their ecology. *Behaviour* **48**: 215–268.
- . 1982. Prospects for interspecific comparison in sociobiology. In *Current problems in sociobiology*, King's College Sociobiology Group (eds.). Cambridge University Press, Cambridge, U.K., p. 323–342.
- , AND M. V. JARMAN. 1979. The dynamics of ungulate social organization. In *Serengeti: Dynamics of an ecosystem*, A. R. E. Sinclair and M. Norton-Griffiths (eds.). The University of Chicago Press, Chicago, Illinois, p. 185–200.
- KEYMER, A. 1981. Population dynamics of *Hymenolepis diminuta* in the intermediate host. *Journal of Animal Ecology* **50**: 941–950.
- LLOYD, M. 1967. Mean crowding. *Journal of Animal Ecology* **36**: 1–30.
- POULIN, R. 1997. Population abundance and sex ratio in dioecious helminth parasites. *Oecologia* **111**: 375–380.
- , AND S. MORAND. 2000. Testes size, body size and male-male competition in acanthocephalan parasites. *Journal of Zoology, London* **250**: 551–558.
- READ, A. F., M. ANWAR, D. SHUTLER, AND S. NEE. 1995. Sex allocation and population structure in malaria and related parasitic protozoa. *Proceedings of the Royal Society of London B* **260**: 359–363.
- , A. NARARA, S. NEE, A. E. KEYMER, AND K. P. DAY. 1992. Gametocyte sex ratios as indirect measures of outcrossing rates in malaria. *Parasitology* **104**: 387–395.
- READ, C. P. 1951. The "crowding effect" in tapeworm infection. *Journal of Parasitology* **37**: 174–178.
- REED, W. J. 1983. Confidence estimation of ecological aggregation indices based on counts—A robust procedure. *Biometrics* **39**: 987–998.
- REICZIGEL, J., AND L. RÓZSA. 2003. Quantitative Parasitology 2.1. Available at: <http://bio.univet.hu/qp/qp.htm>. Accessed on 18 February 2005.
- RÉKÁSI, J., L. RÓZSA, AND J. B. KISS. 1997. Patterns in the distribution of avian lice (Phthiraptera: Amblycera, Ischnocera). *Journal of Avian Biology* **28**: 150–156.
- ROBERTS, S. L. 1961. The influence of population density on patterns and physiology of growth in *Hymenolepis diminuta* (Cestoda: Cyclophyllidae) in the definitive host. *Experimental Parasitology* **11**: 332–371.
- . 2000. The crowding effect revisited. *Journal of Parasitology* **86**: 209–211.
- RÓZSA, L. 1997. Adaptive sex-ratio manipulation in *Pediculus humanus capitis*: Possible interpretation of Buxton's data. *Journal of Parasitology* **83**: 543–544.
- , J. REICZIGEL, AND G. MAJOROS. 2000. Quantifying parasites in samples of hosts. *Journal of Parasitology* **86**: 228–232.
- , J. RÉKÁSI, AND J. REICZIGEL. 1996. Relationship of host coloniality to the population ecology of avian lice (Insecta: Phthiraptera). *Journal of Animal Ecology* **65**: 242–248.
- SASAL, P., E. JOBET, E. FALIEUX, AND S. MORAND. 2000. Sexual competition in an acanthocephalan parasite of fish. *Parasitology* **120**: 65–69.
- SHOSTAK, A. W., AND T. A. DICK. 1987. Individual variability in reproductive success of *Trienophorus crassus* Forel (Cestoda: Pseudophyllidae), with comments on use of the Lorenz curve and Gini coefficient. *Canadian Journal of Zoology* **65**: 2878–2885.
- SZALAY, A. J., AND T. A. DICK. 1989. Differences in numbers and inequalities in mass and fecundity during the egg-producing period for *Raphidascaaris acus* (Nematoda: Anisakidae). *Parasitology* **98**: 489–495.
- ZERVOS, S. 1988a. Population dynamics of a thelastomatid nematode of cockroaches. *Parasitology* **96**: 353–368.
- . 1988b. Evidence for population self-regulation, reproductive competition and arrhenotoky in a thelastomatid nematode of cockroaches. *Parasitology* **96**: 369–379.

### APPENDIX

For the proof of the biasedness of the crowding estimate, let  $I_1, I_2, \dots, I_n$  be an independent random sample taken from a population having no negative and some positive values. Let  $f$  be a monotone increasing function.

The sample mean  $f$ -crowding is defined as

$$\bar{C} = \begin{cases} \frac{\sum_{j=1}^n I_j f(I_j)}{\sum_{j=1}^n I_j}, & \text{for } 0 < \sum_{j=1}^n I_j \\ f(0), & \text{for } 0 = \sum_{j=1}^n I_j, \end{cases}$$

and the population  $f$ -crowding is

$$C = \frac{E[I_1 f(I_1)]}{E(I_1)}.$$

For linear crowding (with no rescaling)  $f(u) = u$  and for logarithmic crowding

$$f(u) = \ln^+(u) = \max(0, \ln(u)).$$

**Theorem:** Suppose that according to the introduced notation the expected values  $E(I_1)$  and  $E(I_1 f(I_1))$  are finite. Then  $E(\bar{C}) \leq C$ , i.e., the bias of the sample mean crowding  $b = E(\bar{C}) - C$  is nonpositive. Furthermore, if  $f(I_1)$  is not constant on the positive part of the population, then the bias is strictly negative.

**Proof:** The sample mean crowding  $\bar{C}$  depends only on the pos-



itive elements of the sample, so its expected value is a weighted average of expected mean crowdings of samples taken from the positive part of the population. More precisely, let  $E_k$  be the expected value of a sample mean crowding of a  $k$ -element sample taken from the positive population, and  $p = P(I_1 > 0)$  the proportion of this part of the population. Then

$$E(\bar{C}) = \sum_{k=0}^n E_k \binom{n}{k} p^k (1-p)^{n-k}.$$

On the other hand,

$$C = \frac{E(I_1 f(I_1) | I_1 > 0)}{E(I_1 | I_1 > 0)},$$

so the population  $f$ -crowding depends also only on the positive population. Therefore, we can restrict the proof of the theorem to populations with positive values only. For  $n = 0$  the bias  $b = f(0) - C$  is negative, because  $f$  increases monotonically. Suppose  $n \geq 1$ . For symmetry, we have and

$$E(\bar{C}) = nE\left(\frac{I_i f(I_i)}{\sum_{j=1}^n I_j}\right) \text{ and } 1 = nE\left(\frac{I_i}{\sum_{j=1}^n I_j}\right),$$

$$i = 1, 2, \dots, n.$$

Hence,

$$b = nE\left\{\frac{I_1}{\sum_{j=1}^n I_j} [f(I_1) - C]\right\}$$

$$= nE\left[E\left\{\frac{I_1}{\sum_{j=1}^n I_j} [f(I_1) - C] | I_2 + I_3 + \dots + I_n\right\}\right].$$

We show that the conditional expectations are nonpositive for all possible conditions such as  $I_2 + I_3 + \dots + I_n = a$  with any fixed  $a$  value. This will prove the theorem. (For  $n = 1$  the condition equals zero, hence it can be omitted.) The bias for layer  $a$  is

$$b_a = nE\left\{\frac{I_1}{I_1 + a} [f(I_1) - C]\right\}.$$

Furthermore,

$$b_a = nE(I_1)E\left\{\left[\frac{1}{I_1 + a} - E\left[\frac{1}{I_1 + a} \cdot \frac{I_1}{E(I_1)}\right]\right] [f(I_1) - C] \frac{I_1}{E(I_1)}\right\}.$$

Let us observe that this equality remains valid if  $C$  is replaced by any other constant  $K$  value. Let

$$g(u) = f\left(\frac{1}{u} - a\right) \text{ and}$$

$$K = g\left\{E\left[\frac{1}{I_1 + a} \cdot \frac{I_1}{E(I_1)}\right]\right\}.$$

Then, we have

$$b_a = nE(I_1)E\left\{\left[\frac{1}{I_1 + a} - E\left[\frac{1}{I_1 + a} \cdot \frac{I_1}{E(I_1)}\right]\right] \times \left[g\left(\frac{1}{I_1 + a}\right) - g\left\{E\left[\frac{1}{I_1 + a} \cdot \frac{I_1}{E(I_1)}\right]\right\}\right] \frac{I_1}{E(I_1)}\right\}.$$

The function  $g$  is monotone decreasing, so the 2 factors to multiply (before and after the symbol  $\times$ ) cannot have the same sign, therefore their product is nonpositive. If  $f(I_1)$  is not constant, then the expression containing this product has a strictly negative expected value. This completes the proof.