esa

ECOSPHERE

CONCEPTS & THEORY

Comparison of measures of crowding, group size, and diversity

Zsolt Lang,[1,]† Lajos Rózsa,[2,3] and Jenő Reiczigel[1]

[1]*Department of Biomathematics and Informatics, University of Veterinary Medicine Budapest, István utca 2, H-1078 Budapest, Hungary*
[2]*MTA-ELTE-MTM Ecology Research Group, Pázmány s. 1/C, H-1117 Budapest, Hungary*
[3]*Evolutionary Systems Research Group, MTA Centre for Ecological Research, Klebelsberg K. u. 3, H-8237 Tihany, Hungary*

Citation: Lang, Z., L. Rózsa, and J. Reiczigel. 2017. Comparison of measures of crowding, group size, and diversity. Ecosphere 8(7):e01897. 10.1002/ecs2.1897

Abstract. In ecology, diversity is often measured as the mean rarity of species in a community. In behavioral sciences and parasitology, mean crowding is the size of the group to which a typical individual belongs. In this paper, focusing mostly on the mathematical aspect, we demonstrate that diversity and crowding are closely related notions. We show that mean crowding can be transformed into diversity and vice versa. Based on this general equivalence rule, notions, relationships, and methods developed in one field can be adapted to the other one. In relation to crowding, we introduce the notion "effective number of groups" that corresponds to the "effective number of species" used in diversity studies. We define new aggregation indices that mirror evenness indices known from diversity theory. We also construct aggregation profiles and orderings of populations based on aggregation indices. By uniting the mathematical interpretation of the ecological notion of diversity and the ethological notion of typical group size (or crowding, in parasitology), our insight opens a new avenue of both theoretical and methodological research. This is exemplified here using real-life abundance data of avian parasites.

Key words: aggregation profile and ordering; crowding; diversity; effective number of groups; evenness.

Received 6 October 2016; revised 23 April 2017; accepted 22 May 2017; final version received 16 June 2017. Corresponding Editor: Laureano A. Gherardi.

† **E-mail:** lang.zsolt@univet.hu

Introduction

For a single group of individuals, crowding is defined by a monotone increasing function of the number of group members, that is, group abundance. Its value is thought to represent the social environment of group members and therefore considered to measure "crowding" from their point of view. Mean crowding in an assemblage of several groups is obtained as the arithmetic mean of crowding values of all individuals. It was first introduced by Lloyd (1967), who suggested both group abundance and group abundance minus one to be the crowding function. Reiczigel et al. (2005) defined identity, logarithmic, and arbitrary monotone increasing crowding functions. Jarman (1974) and Reiczigel et al. (2005, 2008) applied mean crowding to quantify group size from the group members' point of view (as opposed to mean group size, that is, the arithmetic average of group abundances, which is interpreted as group size from the outsiders' point of view).

Considering a multispecies community, the "groups" may be defined as the different species. Mean rarity is commonly applied in such cases to express the diversity of the community. Patil and Taillie (1982) introduced a rarity function of a given species in a community as a monotone decreasing function $R(p)$, where $0 < p \leq 1$ is the proportion of individuals of that species. They defined diversity as the average rarity of species. This family of diversity indices includes a number

of measures frequently met in practice, for example, species richness, Shannon index, Gini-Simpson index, and numerous other ones based on Rényi's concept of generalized entropy (Simpson 1949, Rényi 1961, Shannon and Weaver 1963, Patil and Taillie 1982, Tóthmérész 1995, Lande 1996).

In this paper, we show that the concepts of mean crowding and diversity are closely related. For suitably chosen crowding and rarity functions, mean crowding can be transformed into a diversity index and vice versa. Exploiting this equivalence, notions, relationships, and methods developed in one field can be adapted to the other one. We first introduce the notion of "effective number of groups" and some new aggregation indices that mirror the effective number of species and evenness. Then, we construct aggregation profiles and determine aggregation orderings of populations based on aggregation indices.

As an illustrative example, we analyzed crowding and diversity measures of five louse species parasitizing avian hosts (Rook, *Corvus frugilegus*). Conspecific parasites inhabiting the same host individual constitute a self-evident and natural group. Each bird may host 0–5 different groups representing 0–5 different parasite species.

In Table 1, we summarize the matching notions, concepts, and terminologies of crowding and diversity theory to be used or introduced in this paper.

Methods

Background

Consider a set of N disjoint groups of individuals. Let X_i denote the number of individuals in the ith group, $i = 1, \ldots, N$. Let S be a monotone increasing function defined for positive real numbers. We call S the crowding function. Mean crowding (or subjective group size) is

$$C_S = \sum_{i=1}^{N} X_i \times \frac{S(X_i)}{X.} \quad (1)$$

where $X. = \sum_{i=1}^{N} X_i$ (Reiczigel et al. 2005). From the group members' point of view, C_S is the simple arithmetic mean of crowding values of the individuals' own groups. For example, if S is identity, $N = 3$, $X_1 = 2$, $X_2 = 3$, $X_3 = 1$, then $X. = 6$ and $C_S = (2 + 2 + 3 + 3 + 3 + 1)/6 = 2.33$. Let $p_i = X_i/X.$ denote the proportion of individuals in the ith group. Then, $X_i = X. \times p_i$ and

$$C_S = \sum_{i=1}^{N} p_i \times S(X. \times p_i). \quad (2)$$

If the groups are considered to be the units of observation, then C_S is the weighted mean of transformed group sizes with weights equal to the relative abundances of individuals in the groups. For the previous example, this yields $C_S = 2/6 \times 2 + 3/6 \times 3 + 1/6 \times 1 = 2.33$.

When different groups represent different species, the diversity index of the community is commonly measured by the average rarity of species. Let R be a rarity function. The diversity index according to R is (Patil and Taillie 1982):

$$D_R = \sum_{i=1}^{N} p_i \times R(p_i). \quad (3)$$

Equivalence of diversity and crowding

From Eqs. 2, 3, we see that the structures of the indices of mean crowding and diversity are similar.

Table 1. An overview of parallelism between crowding and diversity concepts, indices, and analysis tools.

Description	A biological population subdivided into groups	An ecological community built of populations representing different species
Origin of concept	Ethology, parasitology	Ecology
Higher-level unit	Population (of similar individuals, e.g., conspecifics)	Ecological community (of different species)
Lower-level units	Groups (e.g., herds, flocks, bands, packs, shoals, or colonies)	Populations (representing different species)
Scale function	Crowding	Rarity
Indices	Mean crowding	Diversity
	Effective number of groups	Effective number of species
	Aggregation	Evenness
Ordering	Aggregation	Diversity
Diagrams	Aggregation profile	Diversity profile

They are related in opposite directions, since the crowding function S is monotone increasing and the rarity function R is monotone decreasing. However, a relevant difference between the characteristics of the two measures is that mean crowding is density-dependent (it is derived from group abundances) and diversity is density-independent (it is calculated from abundance proportions of the groups). To relate them, we keep the total abundance $X_.$ fixed. Under this condition, the function

$$R(p) = S(X_.) - S(p \times X_.) \tag{4}$$

is a rarity function with $R(1) = 0$. The corresponding diversity index is

$$D_R = S(X_.) - C_S. \tag{5}$$

Conversely, starting from an arbitrary rarity function $R(p)$, for any given c constant value

$$S(x) = c - R(x/X_.) \tag{6}$$

is a monotone increasing crowding function, where $0 < x \leq X_.$. If $R(1) = 0$, then $S(X_.) = c$ and Eq. 5 is maintained between the corresponding diversity and crowding indices.

In summary, when the total abundance is held fixed, diversity and crowding indices can be transformed into one another, together with their rarity and crowding functions. We exploit this equivalence to introduce new notions and properties of crowding and diversity.

If the crowding function is the identity function $\mathrm{id}(x) = x$, then mean crowding $C_{\mathrm{id}} = \sum_{i=1}^{N} X_i^2/X_.$ equals Lloyd's (1967) mean demand and linear crowding of Reiczigel et al. (2005). The rarity function assigned to id by Eq. 4 is $R(p) = X_.(1 - p)$, and the diversity index obtained is $X_.$ times the Gini-Simpson diversity $1 - \sum_{i=1}^{N} p_i^2$ (Simpson 1949, Lande 1996, Keylock 2005, Jost 2006). Below, we introduce further pairs of crowding and diversity indices.

Parametric family of crowding and diversity measures

Let

$$S_q(x) = \begin{cases} \dfrac{x^{q-1} - 1}{q - 1}, & \text{if } x > 0 \text{ and } q \geq 0, q \neq 1, \\ \ln(x), & \text{if } x > 0 \text{ and } q = 1 \end{cases} \tag{7}$$

be a family of crowding functions. Here, $\lim_{q \to 1} S_q(x) = \ln(x)$; hence, the family is continuous. Mean crowding indices generated by S_q are

$$C_q = \begin{cases} \dfrac{X_.^{q-1} \times \sum_{i=1}^{N} p_i^q - 1}{q - 1}, & \text{if } q \neq 1 \\ \ln(X_.) + \sum\limits_{i=1}^{N} p_i \times \ln(p_i), & \text{if } q = 1 \end{cases}. \tag{8}$$

The corresponding diversity indices are

$$D_q = \begin{cases} X_.^{q-1} \times \dfrac{1 - \sum_{i=1}^{N} p_i^q}{q - 1}, & \text{if } q \neq 1 \\ -\sum\limits_{i=1}^{N} p_i \times \ln(p_i), & \text{if } q = 1 \end{cases}. \tag{9}$$

Let us observe that D_q can be decomposed into the product of $X_.^{q-1}$ depending only on the total abundance and a purely density-independent factor. The latter is a diversity index related to Rényi's (1961) generalized entropy, studied in detail by Patil and Taillie (1982).

If $q = 2$, then the crowding function is $S_2(x) = x - 1$ and $C_2 = X_. \times \sum_{i=1}^{N} p_i^2 - 1$ is Lloyd's (1967) mean crowding. Its diversity counterpart D_2 is $X_.$ times the Gini-Simpson diversity $1 - \sum_{i=1}^{N} p_i^2$. If $q = 1$, then the crowding function is $S_1(x) = \ln(x)$ and C_1 is the logarithmic crowding (Reiczigel et al. 2005). Its diversity equivalent D_1 is the Shannon diversity index (Shannon and Weaver 1963). When $q = 0$, the crowding function $S_0(x) = 1 - 1/x$ is hyperbolic, so we call

$$C_0 = 1 - \frac{1}{I} \tag{10}$$

the hyperbolic crowding. Here, $I = X_./N^+$ is mean intensity, where N^+ is the number of non-empty groups. (Empty groups may also be interpreted in certain cases, such as when groups of parasites belong to particular host individuals, and non-infected hosts carry "empty groups" of parasites.) The corresponding diversity index is

$$D_0 = X_.^{-1} \times (N^+ - 1). \tag{11}$$

The density-independent factor of D_0 is species richness minus one, $N^+ - 1$.

Both crowding and diversity indices are disproportionally sensitive to large groups for $q > 1$ (e.g., Lloyd's mean crowding and Gini-Simpson diversity) and to small groups for $q < 1$ (e.g., hyperbolic crowding and species richness). The reason is that both C_q and D_q derive from the basic sum $\sum_{i=1}^{N} p_i^q$ when $q \neq 1$. It is also stated

that for $q = 1$, the effect of groups on Shannon's diversity index is proportional to group sizes (Hill 1973, Patil and Taillie 1982, Keylock 2005, Jost 2007). We can carry over this property to logarithmic crowding.

Effective number of groups

A diversity index D generated by a strictly monotone decreasing continuous rarity function R (Eq. 3) is uniquely transformed to an effective number of species N_D by solving (Patil and Taillie 1982)

$$R(N_D^{-1}) = D. \tag{12}$$

Eq. 12 means that a community with N_D equally common (hence equally rare) species has a diversity index equal to D. The quantity called "effective number of species" (numbers equivalent, equivalent number of species, true diversity) was first introduced by MacArthur (1965) in connection with Shannon's diversity index. Hill (1973) developed a family of "diversity numbers" being the effective numbers of species derived from Rényi's entropy (1961). Jost (2006, 2007) argued convincingly for the use of effective number of species to measure diversity, calling it "true diversity."

In a community, the effective number of species usually does not exceed the actual number of species. We formulate the exact rule as an immediate consequence of Theorem 4.3 in Patil and Taillie (1982).

Proposition 1.—Suppose R is a strictly monotone decreasing continuous rarity function and N_D is the corresponding effective number of species. If the function $V(p) = p \times R(p)$ is concave on $0 \le p \le 1$, then we have $0 < N_D \le N^+$, where N^+ is species richness. When V is strictly concave, then equality holds if and only if the species in the community are equally common.

Now we adapt the concept of effective number of species to crowding measures. Let S be a strictly monotone increasing continuous crowding function defined for a population partitioned into N groups. Eq. 4 transforms S into a rarity function R if the total number of individuals $X_\cdot$ is held fixed. We introduce the effective number of groups N_C to be the unique solution of Eq. 12 with diversity $D = D_R$ obtained from Eq. 5. By simple algebra, we reformulate Eq. 12 to get

$$S\left(\frac{X_\cdot}{N_C}\right) = C_S. \tag{13}$$

To visualize the effective number of groups, according to Eq. 13 we can rearrange the total number of individuals $X_\cdot$ into N_C equally abundant groups having the same crowding value C_S. A similar interpretation was presented for the effective number of species, that is, true diversity, in Tuomisto (2010). The unique solution of Eq. 13 is

$$N_C = N \times \frac{\bar{X}}{S^{-1}(C)}. \tag{14}$$

Here, $C = C_S$ is the crowding index related to S and $\bar{X} = X_\cdot/N$ is mean abundance. We can put Eq. 14 in a slightly different form

$$N_C = N^+ \times \frac{I}{S^{-1}(C)} \tag{15}$$

where N^+ is the number of non-empty groups and $I = X_\cdot/N^+$ is mean intensity.

From Eq. 14, we see that the effective number of groups N_C is proportional to the actual number of groups N. The second factor is the ratio of mean abundance $\bar{X}$ to mean crowding transformed back to abundance scale

$$S^{-1}(C) = S^{-1}\left(\frac{\sum_{i=1}^{N} X_i \times S(X_i)}{X_\cdot}\right). \tag{16}$$

We name $S^{-1}(C)$ as the rescaled mean crowding index. It is a Kolmogorov–Nagumo-weighted quasi-arithmetic mean of group abundances (Aczél 1948, Rényi 1961, Beliakov et al. 2007). Being an average, it always falls within the smallest positive and largest group abundance values recorded.

Proposition 1 can be translated to crowding measures based on Eq. 4 linking rarity to crowding functions. Essentially it means that the effective number of groups will not exceed the actual number of groups in the population. In the following form of the statement, we apply the notations used in Eqs. 14, 15.

Proposition 2.—Suppose $X_\cdot > 0$ and the function $W(x) = x \times S(x)$ is convex for $x \ge 0$. Then, we have $0 < N_C \le N^+ \le N$ and $S^{-1}(C_S) \ge I \ge \bar{X} > 0$. When W is strictly convex, equalities hold if and only if the groups in the population are equally sized.

Aggregation and evenness

It follows from Proposition 1 that under general regulatory conditions, the ratio

$$E_R = \frac{N_D}{N^+} \tag{17}$$

is between 0 and 1 and reaches 1 only in the case of perfect evenness of species' proportions. The ratio E_R therefore seems suitable to measure the evenness component of diversity. Hill (1973) was the first to measure it in this manner for the family of parametric diversity indices coinciding with the density-independent part of Eq. 9.

In connection with crowding, the concept of aggregation is preferred instead of evenness (Lloyd 1967, Bez 2000) and, therefore, we introduce a new aggregation index corresponding to crowding function S,

$$A_S = \frac{N}{N_C} = \frac{S^{-1}(C)}{\bar{X}}. \tag{18}$$

To separate the aggregating effects of non-empty groups and their proportion among all groups, we can decompose the aggregation index into the product

$$A_S = \frac{N^+}{N_C} \times \frac{N}{N^+}. \tag{19}$$

Here, the first factor

$$\frac{N^+}{N_C} = \frac{S^{-1}(C)}{I} \tag{20}$$

is the aggregation index of the same form as Eq. 18 calculated for the subset of non-empty groups, and the second factor is the reciprocal of the proportion of non-empty groups (being equal to the prevalence of infested hosts in a host-parasite community).

When the regulatory conditions of Proposition 2 are fulfilled, A_S and its components in Eq. 19 are greater than or equal to 1 and they are equal to 1 only if the groups in the population have the same size. When the crowding function is identity $\mathrm{id}(x) = x$, then the corresponding aggregation index is the ratio of Lloyd's mean demand (Lloyd 1967) or linear crowding presented in Reiczigel et al. (2005) to mean abundance $\bar{X}$

$$A_{\mathrm{id}} = \frac{\frac{1}{N}\sum_{i=1}^{N} X_i^2}{\bar{X}^2}. \tag{21}$$

We note that the linear aggregation index A_{id} is a special case of Bez's aggregation index (Bez 2000). We mention also that Lloyd's patchiness index (Lloyd 1967)

$$P_{\mathrm{Lloyd}} = \frac{C_2}{\bar{X}} = \frac{\frac{1}{N}\sum_{i=1}^{N} X_i \times (X_i - 1)}{\bar{X}^2} \tag{22}$$

is not an aggregation index of the form of Eq. 18. Bez (2000) compared Lloyd's patchiness index with several aggregation indices including A_{id}.

The aggregation indices matching the family of crowding functions S_q in Eq. 7 are

$$A_q = \begin{cases} \left(\dfrac{\frac{1}{N}\sum_{i=1}^{N} X_i^q}{\bar{X}^q} \right)^{\frac{1}{q-1}}, & \text{if } q \geq 0, q \neq 1 \\ \dfrac{\exp\left(\dfrac{\frac{1}{N}\sum_{i=1}^{N} X_i \times \ln(X_i)}{\bar{X}} \right)}{\bar{X}}, & \text{if } q = 1 \end{cases}. \tag{23}$$

They are density-independent (i.e., depend only on relative abundances) and increase strictly monotone with increasing scale parameter q. When $q = 0$, the aggregation index is the reciprocal of the prevalence of non-empty groups. For $q < 1$, the index A_q emphasizes the aggregation among small groups. The sensitivity of the logarithmic aggregation index A_1 is proportional to group abundances (here $q = 1$). When $q > 1$, the index is sensitive to large, dominant groups. If q becomes large, A_q approaches the ratio of the size of the largest group $\max X_i$ to mean abundance $\bar{X}$.

An aggregation profile of a population can be drawn by plotting the natural logarithms of the aggregation indices A_q against the scale parameter q. Aggregation ordering of a family of populations (such as those forming an ecological community) is also possible by comparing their aggregation profiles. The ideas of aggregation profiles and aggregation ordering originate from diversity profiles and diversity ordering described in Patil and Taillie (1982). Tóthmérész (1995) compared different graphical methods to visualize diversity profiles and ordering.

Statistical inference

Our results and formulas described above pertain to assemblages partitioned into a known

number of groups, with fully recorded within-group abundances of individuals. If the target assemblage has a large number of individuals classified in groups, then a random sample is usually drawn to estimate crowding, group size, or diversity characteristics. However, the structure of the groups in the assemblage and the sampling design may affect statistical inference.

Suppose first that each group in the population contains limited number of individuals. Parasites (representing individuals) harbored by hosts (representing groups) is an example of such a population. To estimate crowding measures of the whole population, we take a random sample of size N of the groups. If the number of groups in the population is too small, then groups are selected with replacement to ensure independence. Data of all individuals belonging to the sampled groups are recorded. Let the abundances of individuals in the ith group in the sample be X_i, $i = 1, \ldots, N$. In what follows, we demonstrate that all the crowding and group size measures considered in the previous sections can be consistently estimated from this sample.

Proposition 3.—Let S be a crowding function. Suppose the population averages EX_1 and $EX_1S(X_1)$ are finite and $EX_1 > 0$. Then, the sample estimate (Eq. 1) of the crowding index is consistent, that is, tends to the crowding index of the whole population

$$\frac{EX_1S(X_1)}{EX_1} \tag{24}$$

with probability 1 as the sample size N tends to infinity.

This statement follows from the strong law of large numbers (Etemadi 1981, Gut 2013). The population-level crowding index (Eq. 24) was introduced by Reiczigel et al. (2005). Percentile and bias-corrected and accelerated (BCa) bootstrap confidence intervals (CI) for Eq. 24 can be calculated by resampling (Efron and Tibshirani 1993, Davison and Hinkley 1997, Reiczigel et al. 2005).

Population-level aggregation indices can also be approximated with sample estimates.

Proposition 4.—Suppose S is a strictly monotone continuous crowding function, EX_1 and $EX_1S(X_1)$ are finite, and $EX_1 > 0$. Then, the aggregation index (Eq. 18) estimated from the sample is consistent, that is, tends to the population-level aggregation index

$$\frac{S^{-1}\left(\frac{EX_1S(X_1)}{EX_1}\right)}{EX_1} \tag{25}$$

with probability 1 as the sample size N tends to infinity.

This assertion is also concluded from the strong law of large numbers. Percentile and BCa bootstrap CIs for Eq. 25 can be calculated by resampling (Efron and Tibshirani 1993, Davison and Hinkley 1997).

Consequently, if the moments EX_1^q are positive and finite for all $q > 0$, then sample estimates of the density-independent parametric aggregation indices (Eq. 23) and their aggregation profiles are consistent.

If the total number of groups in the entire population is G, then the population-level effective number of groups is

$$G \times \frac{EX_1}{S^{-1}\left(\frac{EX_1S(X_1)}{EX_1}\right)}. \tag{26}$$

If G is known or its consistent estimate is available, then the effective number of groups is consistently estimated by the ratio of the (estimate of) total number of groups G to the sample aggregation index (Eq. 18).

If the individuals belonging to groups are classified into a limited number of non-overlapping sets (let us call them species), then diversities of species of the whole assemblage can also be estimated. First, a random sample of groups of size N taken with replacement is needed. In the next step, random subsamples of individuals of each group in the sample are drawn independently of each other. Abundance proportions of species of the whole assemblage are estimated based on the obtained subsamples. Finally, diversities of the pooled assemblage are calculated from the proportions of abundances as, for example, in Tuomisto (2010).

Consider a species of the assemblage. Suppose that the abundance of the selected species in the ith sampled group has binomial distribution with size X_i and probability parameter π_i. Let the

mean of probabilities of the selected species in all non-empty groups of the assemblage be π, that is, $E(\pi_1|X_1 > 0) = \pi$. We describe two sampling designs that enable consistent estimates of π and the diversities of the pooled assemblage.

Proposition 5.—Suppose that only non-empty groups are taken in the random sample. Assume that subsamples of individuals of size n are drawn independently with replacement from each sampled group. Let U_i be the subsample abundance of the selected species in group i. Then, the average species proportion $\sum_{i=1}^{N} U_i/(n \times N)$ tends to π with probability 1, when N tends to infinity.

The second sampling design requires that each individual in group i is included in the ith subsample independently of the other individuals with a certain probability of ρ_i. (Therefore, the sizes of the subsamples vary by chance.) Let the assemblage average of the sampling probability $E\rho_1 = \rho > 0$. We mention the important special case $\rho = 1$ when all individuals in the groups are included in the subsamples. See the illustrative analysis of groups of avian lice in the *Example* section.

Proposition 6.—Assume the proportion of non-empty groups $P(X_1 > 0)$ is positive, and sampling probability ρ_i is independent of group size X_i and species probability π_i. Let Z_i be the total abundance and U_i be the abundance of the selected species in the subsample drawn from group i. Then, the average species proportion $\sum_{i=1}^{N} U_i/\sum_{i=1}^{N} Z_i$ tends to $E(\pi_1 \times X_1)/E(X_1)$ with probability 1, when N tends to infinity. If additionally X_i is uncorrelated with π_i for non-empty groups, then the limit equals π.

The proofs of Propositions 5 and 6 are placed in the Proofs section. We mention that the estimates of species proportions based on the sampling design applied in Proposition 6 lead to consistent estimates of gamma diversities (see, e.g., Tuomisto 2010) of the pooled assemblage. When group size X_i and species probability π_i are not correlated in non-empty groups, then the species proportions obtained with the sampling method used in Proposition 5 can also be applied to consistently estimate gamma diversities of the pooled assemblage.

We emphasize that within-group species proportions tend to correlate strongly with group size making comparisons of diversities across groups challenging. This positive association may emerge due to two different causalities. First, we can interpret each group as a random subsample taken from the total flora or fauna. From a purely mathematical point of view, we expect rare species often being absent from smaller groups. The larger the groups are, the more likely they will involve rare species, too. Propositions 5 and 6 control the effects of this mechanism. Second, biogeographical processes (MacArthur and Wilson 1967) may be responsible for a positive correlation between group size X_i and species probability π_i. Say, the sequential colonization of newly emerged habitat patches by different species likely also results in such relationships. Thus, newly born individuals of a host population or new islands emerging in an archipelago are sequentially targeted by new colonizers representing different species.

When the number of groups partitioning the assemblage is limited, we can also choose another data collecting strategy of taking a random sample from the pooled (virtually ungrouped) set of individuals. All available data, including group membership, have to be recorded for all individuals sampled. This method is usually selected when groups contain a great number of individuals, for example, when groups represent species in a community. Another example is the collection of non-overlapping habitat areas, where each area (playing the role of a group) harbors a lot of individuals. In some applications, however, it is possible that the pooled set contains only moderate number of individuals. In this case, the random sample has to be drawn with replacement. Relative abundances of the groups are consistently estimated from samples taken from the pooled assemblage. Estimates of the diversity indices (Eq. 3), the effective numbers of species obtained by solving Eq. 12, the density-independent parametric aggregation indices (Eq. 23), and their aggregation profiles are also consistent. Methods for bias reduction and bootstrap CI estimates together with illustrative simulated and real-world examples are described in Appendix E of Chao et al. (2015), Chao and Jost (2015) for parametric measures estimated from this type of samples.

We mention that it is not possible to estimate density-dependent crowding measures properly,

based merely on the sample drawn from the pooled set of individuals. If an additional estimate of the total abundance X of individuals is available (Seber 1982, Baillargeon and Rivest 2007), then mean crowding C with crowding function S can be estimated by Eq. 2. Furthermore, the effective number of groups

$$G_C = \frac{X}{S^{-1}(C)} \tag{27}$$

is obtained by modifying Eq. 14. If the number of groups G of the whole assemblage is also known or estimated, for example, by rarefaction (Colwell et al. 2012), then the aggregation index derived from Eq. 27 is

$$A_S = \frac{G}{G_C} = \frac{G \times S^{-1}(C)}{X}. \tag{28}$$

Example

To illustrate the notions and methods introduced above, here we present a detail of a statistical analysis of data describing lice parasitizing avian (Rook, *Corvus frugilegus*) hosts. These data were formerly reported by Rózsa et al. (1996). The sample consists of 1973 individuals representing five species of parasitic lice hosted by 37 bird individuals. General descriptive statistics are listed in Table 2.

Three dominant and two rare species comprise the sample. Both prevalence (the proportion of infested hosts within the sample) and mean intensity (the mean number of parasites harbored by infested hosts) are high in the dominant species. Standard deviation (SD) of intensity is greater than mean intensity both within each dominant species and in the pooled louse community, indicating that the abundance distributions are aggregated. We note that both Lloyd's mean crowding and linear crowding are functions of mean and SD of intensity (Lloyd 1967, Reiczigel et al. 2005).

Crowding indices are summarized in Tables 3, 4. Lloyd's and linear crowding indices emphasize louse intensities of highly infected hosts. Their abundance-rescaled versions (Eq. 16) coincide with the linear crowding index. Logarithmic crowding expresses multiplicative crowding effects. It means that a small change (Δ) in its value corresponds to approximately $100 \times \Delta$ percent change in its rescaled equivalent. For instance, the difference between mean logarithmic crowding indices of *Menacanthus gonophaeus* and *Myrsidea isostoma* is $\Delta = 0.15$, producing approximately 15% greater rescaled logarithmic crowding for *Me. gonophaeus*. (The exact difference is $100 \times (\exp(0.15) - 1) = 16.2\%$.) In logarithmic crowding, each host is weighted proportionally to its louse abundance. Therefore, the large difference observed between rescaled mean linear and logarithmic crowding of *My. isostoma* corresponds to the existence of a great proportion of highly infected hosts. In *Brueelia tasniamae*, the difference between rescaled mean linear and logarithmic crowding is moderate indicating a smaller proportion of heavily infected hosts. This pattern can also be seen in the aggregation plot (Fig. 1) where the log-scaled aggregation profiles of *My. isostoma* and *B. tasniamae* intersect.

Hyperbolic crowding indices depend only on mean intensity (Eq. 10). Rescaled hyperbolic mean crowding coincides with mean intensity. The comparison of rescaled linear and logarithmic crowding with mean intensity may reveal

Table 2. General descriptive statistics of louse species harbored by avian hosts.

Species	Total abundance	No. infected hosts	Prevalence (%)	Intensity	
				Mean	SD
Brueelia tasniamae	787	18	48.7	43.7	54.9
Myrsidea isostoma	572	26	70.3	22.0	38.8
Menacanthus gonophaeus	559	24	64.9	23.3	38.5
Philopterus atratus	38	9	24.3	4.2	4.1
Colpocephalum fregili	17	3	8.1	5.7	3.5
All together	1973	34	91.9	58.0	96.7

Notes: Total abundance is the number of parasite individuals, prevalence is the proportion of infested hosts within the sample, and intensity is the number of parasites harbored by an infested host. SD, standard deviation.

Table 3. Parametric crowding indices.

Species	Lloyd		Linear		Logarithmic		Hyperbolic	
	Index	95% BCa CI	Index	95% BCa CI	Index	95% BCa CI	Index	95% BCa CI
Brueelia tasniamae	107.9	79.5, 127.6	108.9	80.5, 128.6	4.49	4.01, 4.75	0.977	0.956, 0.986
Myrsidea isostoma	86.9	24.8, 164.3	87.9	25.8, 165.3	3.89	3.00, 4.83	0.955	0.921, 0.980
Menacanthus gonophaeus	83.4	43.2, 123.0	84.4	44.2, 124.0	4.04	3.28, 4.58	0.957	0.917, 0.978
Philopterus atratus	6.7	2.1, 9.9	7.7	3.1, 10.9	1.81	0.98, 2.30	0.763	0.529, 0.872
Colpocephalum fregili	6.1	1.0, 8.0	7.1	2.0, 9.0	1.88	0.69, 2.20	0.824	0.500, 0.889
All together	213.5	117.9, 350.6	214.5	118.9, 351.6	4.95	4.35, 5.55	0.983	0.970, 0.990

Note: BCa, bias-corrected and accelerated; CI, confidence interval.

aggregation among highly infected and moderately infected hosts. The differences are of the same magnitude as intensity for the three dominant species, indicating heterogeneity of infection, that is, the existence of considerable proportions of highly, moderately, and lowly infected hosts. Contrarily, in case of the two rare species the differences are negligible (Table 4) due to their generally low levels of infection.

We calculated BCa 95% bootstrap CIs for the estimated quantities in Tables 3, 4.

Parametric aggregation indices estimated from relative abundances are shown in Table 5. When all species are pooled together, the linear aggregation ($q = 2$) is 4.0. It is the ratio of mean crowding to mean abundance. It expresses aggregation of parasites on the heavily infected hosts. The interpretation of this aggregation level is simple; "typical" lice live in birds four times more parasitized than "typical" birds. The logarithmic aggregation ($q = 1$), that is, the ratio of rescaled mean logarithmic crowding to mean abundance, is 2.6. It represents the aggregation of lice occurring on moderately parasitized birds. Hyperbolic aggregation ($q = 0$) is the reciprocal of prevalence; its estimated value is 1.1. It reflects only the presence or absence of lice in hosts; it is not influenced by the actual louse abundance values.

In this particular example, the percentile and BCa bootstrap CIs are too wide and, therefore, only the approximate magnitudes of crowding and aggregation measures are estimated safely. We would need a larger collection of abundance records to prove statistically that the differences detected in the sample also characterize the populations. In this illustrative analysis, we emphasize the importance of reporting CIs as a measure of precision of the obtained results. Moreover, CIs can serve as guides in planning appropriate sample sizes of future collections of data.

The profiles of parametric aggregation indices of the five louse species and the pooled community of lice are shown in Fig. 1. Right parts of the curves reflect parasite aggregation on the most infected hosts. The levels of aggregation appear to be similar in case of the three dominant species. The two rare species seem to be much more aggregated, mainly because their aggregation indices are inversely proportional to the low

Table 4. Rescaled parametric crowding indices.

Species	Lloyd		Linear		Logarithmic		Hyperbolic	
	Index	95% BCa CI	Index	95% BCa CI	Index	95% BCa CI	Index	95% BCa CI
Brueelia tasniamae	108.9	80.5, 128.6	108.9	80.5, 128.6	89.1	55.0, 115.7	43.7	22.5, 73.3
Myrsidea isostoma	87.9	25.8, 165.3	87.9	25.8, 165.3	48.8	20.1, 125.6	22.0	12.8, 50.6
Menacanthus gonophaeus	84.4	44.2, 124.0	84.4	44.2, 124.0	56.6	26.8, 98.4	23.3	12.1, 44.9
Philopterus atratus	7.7	3.1, 10.9	7.7	3.1, 10.9	6.1	2.7, 9.9	4.2	2.1, 7.8
Colpocephalum fregili	7.1	2.0, 9.0	7.1	2.0, 9.0	6.5	2.0, 9.0	5.7	2.0, 9.0
All together	214.5	118.9, 351.6	214.5	118.9, 351.6	141.0	77.8, 260.3	58.0	34.1, 104.0

Note: BCa, bias-corrected and accelerated; CI, confidence interval.

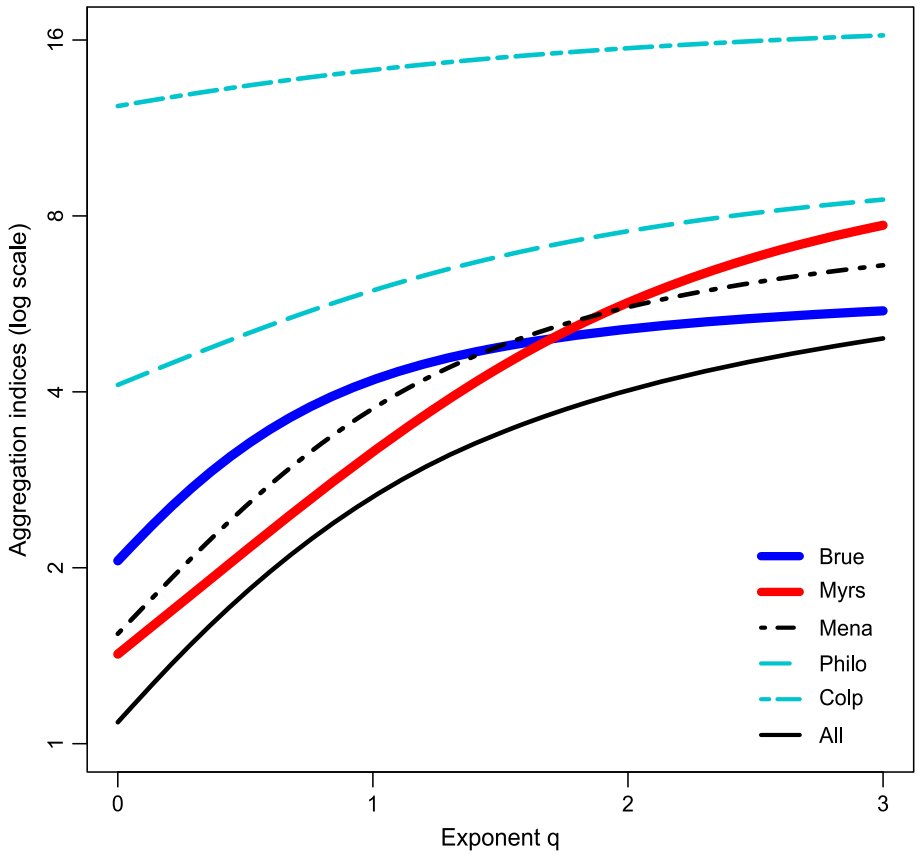

Fig. 1. Profiles of parametric aggregation indices.

Table 5. Parametric aggregation indices.

Species	Exponent	Aggregation index	Percentile CI	BCa CI
Brueelia tasniamae	0	2.1	1.5, 3.1	1.5, 3.1
	1	4.2	2.8, 6.9	2.8, 6.8
	2	5.1	3.2, 9.7	3.1, 9.2
Myrsidea isostoma	0	1.4	1.2, 1.9	1.2, 1.9
	1	3.2	1.8, 4.6	2.1, 5.7
	2	5.7	2.3, 8.4	2.9, 12.5
Menacanthus gonophaeus	0	1.5	1.3, 2.1	1.3, 2.1
	1	3.74	2.5, 5.3	2.7, 5.9
	2	5.6	3.5, 8.9	3.7, 10.0
Philopterus atratus	0	4.1	2.6, 9.3	2.6, 9.3
	1	6.0	3.5, 12.0	3.6, 12.7
	2	7.5	4.2, 15.2	4.4, 16.3
Colpocephalum fregili	0	12.3	5.3, 37.0	5.3, 37.0
	1	14.2	6.3, 37.0	5.2, 37.0
	2	15.5	1.0, 37.0	1.0, 37.0
All species together	0	1.1	1.0, 1.2	1.0, 1.3
	1	2.6	1.9, 3.5	2.1, 3.9
	2	4.0	2.5, 5.9	2.9, 6.9

Note: BCa, bias-corrected and accelerated; CI, confidence interval.

prevalence of infested birds; see Eqs. 19, 20. However, sample size is too small to draw reliable statistical conclusion for these species.

Finally, true diversities of the pooled community of louse species are summarized in Table 6. The Shannon and Simpson true diversities clearly identify the three dominant species. Percentile and BCa 95% bootstrap CIs are quite narrow and accurate.

The R code and data used for the statistical analyses outlined in this section are given in the supporting information (Data S1).

Simulation

To provide an adequate control of the accuracy of the statistical estimates, we simulated data of groups of individuals that contain various species. Group abundances were generated from the equal mixture of two negative binomial distributions having mean value and shape parameters

Table 6. True diversities of the pooled community of louse species.

Name	Exponent	True diversity	Percentile CI	BCa CI
Species richness	0	5.0	4.0, 5.0	4.0, 5.0
Shannon	1	3.3	2.9, 3.6	3.1, 3.7
Simpson	2	3.1	2.5, 3.3	2.7, 3.4

Note: BCa, bias-corrected and accelerated; CI, confidence interval.

200 and 10 and 1000 and 0.5, respectively. Zeroes and the highest 1% of the values of the distributions were excluded. The resulting mixture is heterogeneous and aggregated, with expected mean 572.01, SD 926.53, coefficient of variation (CV) 1.62, and linear aggregation 3.62 (estimated by Eq. 21), producing a wide range of group sizes.

We estimated crowding and aggregation measures (Eqs. 8, 16, 23) together with percentages of bias and standard error (SE; related to assemblage-level means) for exponents $q = 0, 1, 2$. Propositions 3 and 4 demonstrate that the estimates are consistent. Sample sizes of group abundances were set to 50, 100, 300, 500, and 1000. We obtained the estimates from 10^5 simulated replicates of group abundance samples for each sample size. Assemblage-level crowding and aggregation measures were calculated from 10^7 simulated group abundances. They were used to assess bias of the estimates. The results are summarized in Table 7.

The estimates were nearly unbiased for $q = 0$ and had a moderate negative bias for $q = 1, 2$. For crowding indices, this was proven mathematically in Reiczigel et al. (2005). When the sample sizes were at least 100, bias was $<2.5\%$ and SE was 10–21% of the assemblage mean for $q = 1, 2$. When the sample sizes were 300, bias was $<0.75\%$ and SE was 5–12% of the assemblage mean for $q = 1, 2$. Reiczigel et al. (2005) obtained roughly the same level of accuracy for empirical distributions of avian ectoparasites.

Within-group theoretical species proportions were derived from the lognormal (Magurran 2004) and geometric series (Tokeshi 1990) distribution models. To account for between-group heterogeneity, the actual within-group species proportions were generated (independently of the abundance of the group) from a Dirichlet distribution (McCarthy 2007) with parameter vector proportional to the theoretical species proportions. The Dirichlet model ensures that the expected (assemblage mean) within-group species proportions coincide with the predefined theoretical species proportions. Within-group species abundances were simulated from the multinomial distribution with size being the abundance of the group and probabilities equal to the actual within-group species proportions.

Table 7. Mean, bias, and standard error (SE) of estimates of parametric crowding and aggregation measures.

		Crowding index			Rescaled crowding index			Aggregation index		
Exponent	No. groups	Estimate	Bias (%)	SE (%)	Estimate	Bias (%)	SE (%)	Estimate	Bias (%)	SE (%)
0	50	0.998	−0.01	0.04	569.990	0.00	22.94	1.000	0.00	0.00
	100	0.998	−0.01	0.03	570.985	0.00	16.17	1.000	0.00	0.00
	300	0.998	0.00	0.02	572.316	0.00	9.35	1.000	0.00	0.00
	500	0.998	0.00	0.01	572.031	0.00	7.14	1.000	0.00	0.00
	1000	0.998	0.00	0.01	572.227	0.00	5.13	1.000	0.00	0.00
1	50	7.062	−0.97	4.33	1221.220	−2.31	28.79	2.119	−3.02	12.34
	100	7.098	−0.47	3.02	1236.699	−1.07	20.68	2.154	−1.42	8.82
	300	7.121	−0.14	1.69	1246.911	−0.26	11.94	2.175	−0.47	4.98
	500	7.125	−0.09	1.30	1247.689	−0.19	9.19	2.179	−0.29	3.84
	1000	7.128	−0.04	0.92	1249.406	−0.06	6.55	2.182	−0.14	2.71
2	50	1958.405	−5.47	27.81	1959.405	−5.47	27.80	3.431	−5.32	18.63
	100	2020.101	−2.49	20.13	2021.101	−2.49	20.12	3.539	−2.33	13.33
	300	2056.569	−0.73	11.49	2057.569	−0.73	11.49	3.597	−0.74	7.55
	500	2061.855	−0.48	8.88	2062.855	−0.48	8.87	3.607	−0.46	5.80
	1000	2067.684	−0.20	6.27	2068.684	−0.20	6.27	3.616	−0.22	4.10

The number of species was set to 200 in the simulated assemblage. In the lognormal model, the kth theoretical species proportion p_k was set c times the k/201-quantile of the lognormal distribution with location parameter 0 and scale parameter 1.2. In the geometric series model, $p_k = c \times 0.03 \times 0.97^{k-1}$, $k = 1, 2, \ldots, 200$. In each version, c is a normalizing constant to ensure that the sums of the theoretical species proportions equal 1. The CV of both models was 1.43, so their dominant species proportions were equally heterogeneous. For moderately frequent and rare species, the lognormal model is more diverse and even than the geometric series model; their Shannon diversities are 106.74 and 87.82, respectively. The CV of the actual within-group species proportions generated from the Dirichlet distribution was 1.79; it is 25% greater than the CV of the theoretical species proportions.

We estimated assemblage-level gamma diversities together with their percentage bias and SE for exponents $q = 0, 1, 2$. In a single simulation cycle, the number of groups was set to 20, 50, and 100. The sizes of within-group samples of individuals taken with replacement were 20, 50, 100, and 200. The estimates were obtained from 10^5 replicates for each combination of selected group numbers and within-group sample sizes. Exact assemblage-level gamma diversities were calculated from the theoretical species proportion models. They were used to assess bias of the estimates. (We note that according to Proposition 5, the estimates are consistent.) The results are summarized in Table 8.

The simulation demonstrated the existence of considerable negative bias exceeding in most cases the magnitude of SE.

For $q = 0$, the theoretical species richness of 200 was estimated. For small and moderate sample sizes, the observed bias was much larger than the SE. In the assemblage with lognormal species distribution, at least 2000 individuals had to be sampled (100 groups, 20 individuals per group) to achieve $<10\%$ bias. (The corresponding SE was 2%.) The assemblage generated from the geometric series model had much more uneven species distribution. Here, 10,000 individuals were needed (100 groups, 100 individuals per group) to produce 10.7% bias and 1.8% SE.

For $q = 1$, the assemblage-level Shannon diversities were estimated. Their values in the lognormal and geometric series models were 106.74 and 87.82, respectively. The observed bias and SE were generally smaller than those recorded for species richness. The accuracies of the diversity estimates of the two species distribution models were similar. In the logarithmic model with 100 groups sampled and 20 individuals drawn within each group, the recorded bias and SE were −5.6% and 2.9%, respectively. In the geometric series model with the same sampling design, the bias was −6.4% and the SE was 2.3%.

For $q = 2$, the assemblage-level Simpson diversities of the lognormal and geometric series models were 65.61 and 65.37, respectively. They are almost equal, because the CVs of the two species proportion models were the same (up to rounding error). The observed bias was somewhat smaller and the SE was slightly larger than that for Shannon's diversity. The accuracies of the diversity estimates of the two species distribution models were similar. In the logarithmic model, the bias was slightly smaller and the SE larger than that in the geometric model. In the logarithmic model, if 100 groups were drawn and 20 individuals were sampled within each group, then the bias was −3.1% and the SE was 4.8%. In the geometric series model with the same sampling design, the bias was −4.6% and the SE was 2.9%.

We also carried out a modified simulation in a way that within each group each individual was randomized, independently of each other and the actual group size, to be included in the sample with a probability of 0.1. Proposition 6 ensures consistency of the estimates of assemblage-level gamma diversities obtained from this sampling design. In the logarithmic model with 50 groups, for $q = 0$ the bias and SE of the estimated gamma diversity were −9.1% and 2.6%, respectively. For $q = 1$, the bias and SE were −7.1% and 2.9%, and for $q = 2$, they were −4.5% and 4.9%, respectively. In the geometric series model with 50 groups, for $q = 0$ the bias and SE of the estimated gamma diversity were −25.3% and 3.3%, respectively. For $q = 1$, the bias and SE were −6.5% and 2.5%, and for $q = 2$, they were −4.7% and 3.1%, respectively.

When group abundances and the number of species in the assemblage are small, it is convenient to draw all individuals (without replacement) of the sampled groups to estimate

Table 8. Mean, bias, and standard error (SE) of estimates of parametric diversities.

Exponent	No. groups	Within-group sample size	Diversities of the lognormal model			Diversities of the geometric series model		
			Estimate	Bias (%)	SE (%)	Estimate	Bias (%)	SE (%)
0	20	20	118.63	−40.69	2.88	96.09	−51.96	2.26
		50	153.42	−23.29	3.13	120.54	−39.73	2.35
		100	171.37	−14.32	3.05	135.76	−32.12	2.40
		200	182.07	−8.97	2.81	147.66	−26.17	2.40
	50	20	159.04	−20.48	2.59	124.94	−37.53	2.28
		50	182.36	−8.82	2.15	148.00	−26.00	2.18
		100	191.46	−4.27	1.68	161.99	−19.00	2.10
		200	195.71	−2.14	1.26	172.03	−13.98	1.97
	100	20	180.57	−9.72	1.97	145.69	−27.15	2.17
		50	193.74	−3.13	1.29	166.90	−16.55	1.96
		100	197.60	−1.20	0.84	178.65	−10.67	1.77
		200	199.06	−0.47	0.53	186.42	−6.79	1.54
1	20	20	79.53	−25.49	4.92	68.32	−22.21	4.15
		50	90.98	−14.76	4.75	75.82	−13.66	3.69
		100	95.60	−10.44	4.59	78.79	−10.28	3.44
		200	97.95	−8.23	4.59	80.41	−8.44	3.27
	50	20	94.39	−11.57	3.83	78.10	−11.07	3.07
		50	100.34	−6.00	3.26	82.00	−6.63	2.45
		100	102.45	−4.03	3.01	83.49	−4.93	2.19
		200	103.52	−3.02	2.86	84.28	−4.04	2.03
	100	20	100.74	−5.63	2.86	82.22	−6.38	2.33
		50	103.97	−2.60	2.37	84.59	−3.69	1.80
		100	105.00	−1.63	2.15	85.37	−2.79	1.60
		200	105.54	−1.13	2.01	85.82	−2.28	1.45
2	20	20	54.84	−16.42	8.36	53.69	−17.87	5.67
		50	59.30	−9.63	7.50	58.15	−11.04	5.18
		100	61.01	−7.02	6.99	59.82	−8.49	4.94
		200	61.84	−5.76	6.83	60.68	−7.18	4.73
	50	20	61.15	−6.80	6.37	60.03	−8.18	3.97
		50	63.28	−3.55	5.24	62.09	−5.01	3.25
		100	64.01	−2.45	4.67	62.85	−3.86	2.93
		200	64.42	−1.81	4.32	63.23	−3.27	2.77
	100	20	63.58	−3.10	4.78	62.39	−4.55	2.92
		50	64.72	−1.36	3.80	63.59	−2.73	2.32
		100	65.10	−0.78	3.39	63.93	−2.20	2.09
		200	65.30	−0.48	3.12	64.14	−1.88	1.93

Note: Within-group individuals were sampled with replacement.

assemblage-level gamma diversities (see the *Example* section for illustration). It follows from Proposition 6 that these estimates are consistent. We investigated the accuracy of the estimates by simulation.

For group abundances, we selected a negative binomial distribution with mean value 60 and shape parameter 0.15, truncated to have only positive values. This is an aggregated distribution typical for abundances of avian ectoparasites, with expected mean 101.17, SD 190.80, CV 1.89, and linear aggregation index 4.56.

The number of species was set to 20 in the simulated assemblage. We applied lognormal and geometric series models to define theoretical species proportions. In the lognormal model, the kth theoretical species proportion p_k was obtained from the k/21-quantiles of the lognormal distribution with location parameter 0 and scale parameter 1.6. In the geometric series model, p_k was proportional to $0.264 \times 0.736^{k-1}$, $k = 1, 2, \ldots, 20$. The CV of both models was 1.43, assuring equal heterogeneity of the dominant species proportions. The actual species proportions and abundances were

generated from the Dirichlet and multinomial distribution the same way as in the previous simulation designs. The CV of the actual within-group species proportions was 1.79; it is 25% greater than the CV of the theoretical species proportions.

Assemblage-level gamma diversities together with their bias and SE were estimated for exponents $q = 0, 1, 2$. In each simulation cycle, the number of groups generated was set to 20, 50, and 100. The estimates were obtained from 10^5 replicates. Exact assemblage-level gamma diversities were calculated from the theoretical species proportion models. They were used to assess bias of the estimates. The results are summarized in Table 9.

The simulation demonstrated the existence of small-to-moderate negative bias and SE that rapidly attenuated when the number of sampled groups increased.

For $q = 0$, the theoretical species richness of 20 was estimated. In the logarithmic model for 20 sampled groups, the resulting bias and SE were −6.5% and 4.9%, respectively. In the geometric series model, drawing 20 groups yielded −16.8% bias and 6.8% SE. In both models with 50 or more sampled groups, the magnitude of bias and SE was less than 6.8% and 4.9%, respectively. For $q = 1$, the assemblage-level Shannon diversities were estimated. In both models with 50 or more sampled groups, the magnitude of observed bias and SE was <4.0% and 6.8%, respectively. For $q = 2$, the Simpson diversities were estimated. In both models with 50 or more sampled groups, the magnitude of observed bias and SE was <2.4% and 9.9%, respectively.

The R code to generate the simulated data used in this section is given in the supporting information (Data S2).

Discussion

In ecology, crowding expresses the magnitude of coexisting individuals in a close neighborhood (individuals belonging to a particular group or inhabiting a habitat patch) of each individual in a population. Diversity is associated with the number of species and their degree of evenness in a community. Although their underlying ecological concepts and application purposes seem markedly different, mean crowding and diversity measures are surprisingly similar in mathematical structure. Both summarize heterogeneity of hierarchical communities of groups of individuals. They are weighted arithmetic averages of generally nonlinear functions of group sizes named crowding function and rarity function. In this paper, we constructed transforming equations between the two families of these measures. The obtained equivalence allowed us to introduce several notions into one field that had been developed historically in the other one.

Crowding functions are usually density-dependent functions of group abundance, total biomass, plant coverage, lifespan, etc. (For the sake of simplicity, we mentioned only abundance throughout the present text.) Rarity functions chosen in diversity index definitions depend on relative abundance, proportion of biomass, etc., so they are density-independent in general. However, the transformation equations (Eqs. 4, 6) can be applied to density-independent

Table 9. Mean, bias, and standard error (SE) of estimates of parametric diversities.

		Diversities of the lognormal model			Diversities of the geometric series model		
Exponent	No. groups	Estimate	Bias (%)	SE (%)	Estimate	Bias (%)	SE (%)
0	20	18.70	−6.48	4.85	16.64	−16.80	6.76
	50	19.74	−1.30	2.38	18.64	−6.79	4.92
	100	19.97	−0.15	0.85	19.56	−2.21	3.05
1	20	8.95	−8.10	9.43	8.07	−7.98	8.56
	50	9.35	−3.96	6.79	8.42	−3.93	6.15
	100	9.54	−2.05	5.14	8.57	−2.28	4.61
2	20	6.30	−4.20	13.77	6.19	−5.45	11.21
	50	6.45	−1.94	9.85	6.39	−2.36	7.91
	100	6.52	−0.81	7.36	6.46	−1.37	5.81

Note: All within-group individuals were drawn in the samples.

crowding functions or density-dependent rarity functions as well. Furthermore, the relationships (Eqs. 4, 6) can be interpreted either to be density-independent under the condition of fixed total abundance or unconditionally, depending also on the varying total abundance.

Mean crowding measures of a population (that is subdivided into groups of individuals) are affected by the crowding function. Its selection is determined by the underlying biological model. The simplest and most frequently applied version is the original abundance scale yielding linear mean crowding, or abundance minus one to produce Lloyd's mean crowding. These scales, however, overweight dominant groups of the population. When the crowding function is logarithmic, each group is weighted proportionally to its abundance in mean crowding. Logarithmic scale seems suitable when there are large differences between group abundances, such as in case of microbial pathogens inhabiting host individuals as habitat patches. Log transformation contracts group abundances of different magnitudes to enable comparisons. Furthermore, logarithmic scale is applicable when the crowding model is multiplicative, that is, when the degree of crowding effect is properly expressed in percentages. These crowding functions are special cases of a family of power expressions (Eq. 7) originating from diversity index theory (Hill 1973, Patil and Taillie 1982).

We note that there are important alternative models beyond family (Eq. 7). For instance, an exponentially increasing crowding function describes well such grouped populations where each new individual entering a group strengthens within-group competition substantially.

A substantive drawback of using different crowding functions is that mean crowding or group size measures obtained in this way are on different scales and therefore cannot be simply compared. Abundance-rescaled crowding indices (Eq. 16), however, have the same scale. They are comparable even when the crowding functions differ. The general aggregation index (Eq. 18) defined as the ratio of rescaled crowding index to mean abundance is also comparable between different crowding scales. This provides justification for the use of aggregation profiles and ordering.

Crowding measures are usually applied to populations containing a great number of disjoint groups each of them consisting of a limited number of individuals. Contrarily, diversity indices are generally designed for communities with bounded number of species that may contain very large numbers of individuals. These minor structural differences are reflected in the methodology of statistical inference. Crowding measures are usually estimated from samples containing entire groups of individuals drawn, if necessary, with replacement. Each sample collected in this way has a set of independent abundance values. On the other hand, a diversity measure is most frequently estimated from an independent sample of individuals of the community investigated. Here, group abundances are not independent for they add up to the sample size, that is, the number of individuals in the sample. Statistical methods have to be chosen according to the different properties of the two sampling schemes.

Proofs

Proof of Proposition 5.—According to the strong law of large numbers, the average species proportion $\sum_{i=1}^{N} U_i/(n \times N)$ tends to $E(U_1|X_1 > 0)/n$ with probability 1, when N tends to infinity. Let Z_i be the abundance of the selected species in group i. We have

$$\begin{aligned} E(U_1|X_1 > 0) &= E(E(U_1|Z_1, X_1, X_1 > 0)) \\ &= E\left(n \times \frac{Z_1}{X_1}|X_1 > 0\right) \\ &= n \times E\left(\frac{Z_1}{X_1}|X_1 > 0\right). \end{aligned}$$

Here,

$$\begin{aligned} E\left(\frac{Z_1}{X_1}|X_1 > 0\right) &= E\left(E\left(\frac{Z_1}{X_1}|\pi_1, X_1, X_1 > 0\right)\right) \\ &= E\left(\frac{X_1 \times \pi_1}{X_1}|X_1 > 0\right) \\ &= E(\pi_1|X_1 > 0) = \pi. \end{aligned}$$

Proof of Proposition 6.—We conclude from the strong law of large numbers that the average species proportion $\sum_{i=1}^{N} U_i/\sum_{i=1}^{N} Z_i$ tends to $E(U_1)/E(Z_1)$ with probability 1, when N tends to infinity. We have

$$E(Z_1) = E(E(Z_1|\rho_1, X_1)) = E(\rho_1 \times X_1) = E(\rho_1) \times E(X_1) \quad \text{(P1)}$$

and

$$E(U_1) = E(E(U_1|\rho_1, \pi_1, X_1)) = E(\rho_1 \times \pi_1 \times X_1) = E(\rho_1) \times E(\pi_1 \times X_1). \quad \text{(P2)}$$

From Eqs. P1, P2, we have

$$\frac{E(U_1)}{E(Z_1)} = \frac{E(\pi_1 \times X_1)}{E(X_1)}.$$

If X_i and π_i are not correlated in non-empty groups, then

$$\frac{E(\pi_1 \times X_1)}{E(X_1)} = \frac{E(\pi_1 \times X_1|X_1 > 0)}{E(X_1|X_1 > 0)} = E(\pi_1|X_1 > 0) = \pi.$$

Acknowledgments

We are grateful to János Podani for valuable comments and suggestions that improved the manuscript. Thanks to József Rékási who collected most of the rook lice utilized in the present paper. We thank the editor (Laureano Gherardi) and an anonymous reviewer for their valuable suggestions that greatly improved this paper. This work was supported by the National Scientific Research Fund of Hungary (OTKA grant no. 108571), the grant "In the light of evolution: theories and solutions" (GINOP-2.3.2-15-2016-00057), and the Research Faculty Grant 2015 of the Szent István University, Faculty of Veterinary Science.

Literature Cited

Aczél, J. 1948. On mean values. Bulletin of the American Mathematical Society 54:392–400.

Baillargeon, S., and L. P. Rivest. 2007. Rcapture: loglinear models for capture-recapture in R. Journal of Statistical Software 19:1–31.

Beliakov, G., A. Pradera, and T. Calvo. 2007. Averaging functions. Pages 39–122 *in* G. Beliakov, A. Pradera, and T. Calvo, editors. Aggregation functions: a guide for practitioners. Springer-Verlag, Berlin, Germany.

Bez, N. 2000. On the use of Lloyd's index of patchiness. Fisheries Oceanography 9:372–376.

Chao, A., T. C. Hsieh, R. L. Chazdon, R. K. Colwell, and N. J. Gotelli. 2015. Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. Ecology 96:1189–1201.

Chao, A., and L. Jost. 2015. Estimating diversity and entropy profiles via discovery rates of new species. Methods in Ecology and Evolution 6:873–882.

Colwell, R. K., A. Chao, N. J. Gotelli, S. Y. Lin, C. X. Mao, R. L. Chazdon, and J. T. Longino. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. Journal of Plant Ecology 5:3–21.

Davison, A. C., and D. V. Hinkley. 1997. Bootstrap methods and their application. Cambridge University Press, Cambridge, UK.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman & Hall, New York, New York, USA.

Etemadi, N. 1981. An elementary proof of the strong law of large numbers. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 55: 119–122.

Gut, A. 2013. The law of large numbers. Pages 265–328 *in* A. Gut, editor. Probability: a graduate course. Springer, New York, New York, USA.

Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. Ecology 54:427–432.

Jarman, P. J. 1974. The social organisation of antelope in relation to their ecology. Behaviour 48:215–268.

Jost, L. 2006. Entropy and diversity. Oikos 113:363–375.

Jost, L. 2007. Partitioning diversity into independent alpha and beta components. Ecology 88:2427–2439.

Keylock, C. 2005. Simpson diversity and Shannon-Wiener index as special cases of a generalized entropy. Oikos 109:203–207.

Lande, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos 76:5–13.

Lloyd, M. 1967. Mean crowding. Journal of Animal Ecology 36:1–30.

MacArthur, R. H. 1965. Patterns of species diversity. Biological Reviews 40:510–533.

MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Monographs in population biology I. Princeton University Press, Princeton, New Jersey, USA.

Magurran, A. E. 2004. Measuring biological diversity. Blackwell, Oxford, UK.

McCarthy, M. A. 2007. Bayesian methods for ecology. Cambridge University Press, Cambridge, UK.

Patil, G. P., and C. Taillie. 1982. Diversity as a concept and its measurement. Journal of the American Statistical Association 77:548–561.

Reiczigel, J., Z. Lang, L. Rózsa, and B. Tóthmérész. 2005. Properties of crowding indices and statistical tools to analyse crowding data. Journal of Parasitology 91:245–252.

Reiczigel, J., Z. Lang, L. Rózsa, and B. Tóthmérész. 2008. Measures of sociality: two different views of group size. Animal Behaviour 75:715–721.

Rényi, A. 1961. On measures of entropy and information. Pages 547–561 *in* J. Neyman, editor. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Volume 1. University of California Press, Berkeley, California, USA.

Rózsa, L., 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.

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. Macmillan, New York, New York, USA.

Shannon, C. E., and W. Weaver. 1963. The mathematical theory of communication. University of Illinois Press, Urbana, Illinois, USA.

Simpson, E. H. 1949. Measurement of diversity. Nature 163:688.

Tokeshi, M. 1990. Niche apportionment or random assortment: species abundance patterns revisited. Journal of Animal Ecology 59:1129–1146.

Tóthmérész, B. 1995. Comparison of different methods for diversity ordering. Journal of Vegetation Science 6:283–290.

Tuomisto, H. 2010. A consistent terminology for quantifying species diversity? Yes, it does exist. Oecologia 164:853–860.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1897/full