



The zero-sum assumption in neutral biodiversity theory

Rampal S. Etienne^{a,*}, David Alonso^b, Alan J. McKane^c

^a*Community and Conservation Ecology Group, Centre for Ecological and Evolutionary Studies, University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands*

^b*Ecology and Evolutionary Biology, University of Michigan, 830 North University Av, Ann Arbor, MI 48109-1048, USA*

^c*Theory Group, School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK*

Received 8 March 2007; received in revised form 24 May 2007; accepted 6 June 2007

Abstract

The neutral theory of biodiversity as put forward by Hubbell in his 2001 monograph has received much criticism for its unrealistic simplifying assumptions. These are the assumptions of functional equivalence among different species (neutrality), the assumption of point mutation speciation, and the assumption that resources are continuously saturated, such that constant resource availability implies constant community size (zero-sum assumption). Here we focus on the zero-sum assumption. We present a general theory for calculating the probability of observing a particular species-abundance distribution (sampling formula) and show that zero-sum and non-zero-sum formulations of neutral theory have exactly the same sampling formula when the community is in equilibrium. Moreover, for the non-zero-sum community the sampling formula has this same form, even out of equilibrium. Therefore, the term “zero-sum multinomial (ZSM)” to describe species abundance patterns, as coined by Hubbell [2001. *The Unified Neutral Theory of Biodiversity and Biogeography*, Princeton University Press, Princeton, NJ], is not really appropriate, as it also applies to non-zero-sum communities. Instead we propose the term “dispersal-limited multinomial (DLM)”, thus making explicit one of the most important contributions of neutral community theory, the emphasis on dispersal limitation as a dominant factor in determining species abundances.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Biodiversity; Neutral model; Species abundances; Allele frequencies; Ewens sampling formula; Etienne sampling formula

1. Introduction

Biodiversity is not only determined by the number of different entities (species in community ecology or alleles in population genetics), but also by the abundance of these entities (Magurran, 2004). Traditionally, species abundances have been mostly studied for competitive communities (Hubbell, 2001). Recently, they have also been heralded as important determinants in food webs (Cohen et al., 2003). Abundance data are relatively easy to collect, particularly in community ecology, and thus are good candidates to provide information on the processes that determine biodiversity. For many decades models have been proposed to describe and explain observed patterns in species abundances or allele frequencies (Fisher et al., 1943;

Preston, 1948, 1962; MacArthur, 1957, 1960; Pielou, 1969, 1975; Ewens, 1972; Sugihara, 1980; Tokeshi, 1990, 1993, 1996; Engen and Lande, 1996a,b; Diserud and Engen, 2000; Dewdney, 1998, 2000; Hubbell, 2001; Harte et al., 1999; Etienne and Olf, 2005), but consensus on a single adequate model has not been reached. This has even led some scientists to claim that abundance data cannot distinguish between different models (Volkov et al., 2005; but see Etienne et al., 2007). Although it is true that patterns never uniquely imply process (Cohen, 1968; Clinchy et al., 2002; Purves and Pacala, 2005), scientific progress can still be made by testing whether different hypotheses on plausible processes can predict observed patterns. As for all scientific tests, no theory can be proved to be true, but inadequate theories can certainly be rejected. However, in comparisons of alternative abundance models, a solid sampling theory that provides the likelihoods of these models given the data (i.e. sampling

*Corresponding author. Tel.: +31 503 63 2230.

E-mail address: r.s.etienne@rug.nl (R.S. Etienne).

formulas) has been lacking (Chave et al., 2006). In this paper, we present such a sampling theory. For convenience we will speak only of species abundances in an ecological community, but our results also apply to allele abundances (frequencies) in a population.

Our sampling theory is inspired by neutral theory, as this theory provides a null model for abundance distributions, but we stress that the theory extends beyond neutral theory. Nevertheless, we use our sampling theory here to settle a debate in neutral theory concerning the zero-sum assumption (Hubbell, 2001). Neutral theory has three basic assumptions: the neutrality assumption (functional equivalence among different species), the point mutation assumption (speciation takes place by point mutation) and the zero-sum assumption, which have all received much criticism. Here we do not enter the debate concerning the first two assumptions, but focus on the third. The weak formulation of the zero-sum assumption states that individuals of different species in a community that are limited by their environment (either because of limited shared resource, e.g. space or light, or because of natural enemies that restrain the otherwise unbridled growth of the species) will always saturate the environment, i.e. there are never fewer individuals than allowed by the environment. The strong formulation states that the limitation by the environment (e.g. the number of resources or the number of natural enemies) is constant and that therefore the total number of individuals competing for the resource is also constant. While there is some evidence for this assumption (Hubbell, 1979, 2001), this assumption (in either form) has been felt to be too restrictive (Volkov et al., 2003; Poulin, 2004). Here we show, however, that the zero-sum assumption is not crucial for one of the most important predictions of the theory, the sampling formula, i.e. the formula that gives the probability of observing abundances n_1, \dots, n_S in a sample of J individuals. This is not trivial, as the characteristic time scales of zero-sum and non-zero-sum models are very different (see below).

We will first present our general sampling theory, particularly its fundamental assumptions and the general sampling formula based on these assumptions. We will then formulate a general master equation for a single species and show how it can lead to a general sampling formula for multiple neutral species that have either fully independent (non-zero-sum) or fully dependent (zero-sum) dynamics. For the latter case we will introduce the subsample approach. Up to this point, our treatment will be quite general in the sense that it does not rely on any assumptions concerning the dynamics of each species such as intraspecific competition, or the source of new species such as speciation or immigration. From this point we will proceed by inserting such details of the Hubbell model for the metacommunity and for the local community and we will subsequently link the two community scales. We do this for the non-zero-sum case both in and out of equilibrium and refer to the Appendix for the zero-sum equilibrium case. We will demonstrate that the sampling

formula for the zero-sum case in equilibrium also applies to the non-zero-sum case, both in equilibrium and out of equilibrium. We end with a discussion of our results. For the reader's convenience the mathematical symbols used in this paper are summarized in Table 1.

2. A sampling theory

The metacommunity concept represents a powerful framework to study the relative importance of local species interactions and dispersal in determining community composition and dynamics (Leibold et al., 2004; Holyoak et al., 2005). Our sampling theory is based on a classical metacommunity structure. First, we assume that local communities form a metacommunity by global migration (global dispersal, propagule mixing and propagule rain (see Fig. 1)). Such a migration assumption has a long tradition in metapopulation ecology (Gotelli and Kelley, 1993; Bascompte et al., 2002), and has been used to gain insight into the processes underlying the distribution of species diversity from local to global spatio-temporal scales (MacArthur and Wilson, 1967), has inspired current neutral theory (Hubbell, 2001), and underlies recent studies on community similarity under neutrality (Dornelas et al., 2006). Second, as in Hubbell's (2001) approach, our theory builds on the assumption that ecological and evolutionary time scales are very different and can be decoupled. Under these assumptions, our general framework requires the following three general steps:

1. We assume that ecological and evolutionary processes have determined metacommunity composition (species abundances) at the largest spatio-temporal scale. Neutral speciation, adaptive speciation and trade-off invariance are possible mechanisms. We usually evaluate the stationary distribution of species abundances emerging from these processes. We do not require that this distribution is frozen-stable. We do require, however, that processes at the biogeographic level are at a much longer temporal scale relative to the process affecting the ecological assembly of local communities. (See Vallade and Houchmandzadeh, 2006 for a relaxation of this assumption.)
2. In local communities (or islands) mainly local ecological processes, but not speciation, are at play. Examples of these processes are dispersal-limitation, density dependence, habitat heterogeneity and species differential adaptation to different habitats.
3. Sampling formulas contain information about both levels of description and about the sampling process. They enable us to empirically evaluate the imprint of evolutionary processes at the biogeographic scale and ecological processes at the local scale on the observed diversity in our local samples.

When sampling communities, the probability of obtaining S species with n_1, \dots, n_S individuals in a sample of size

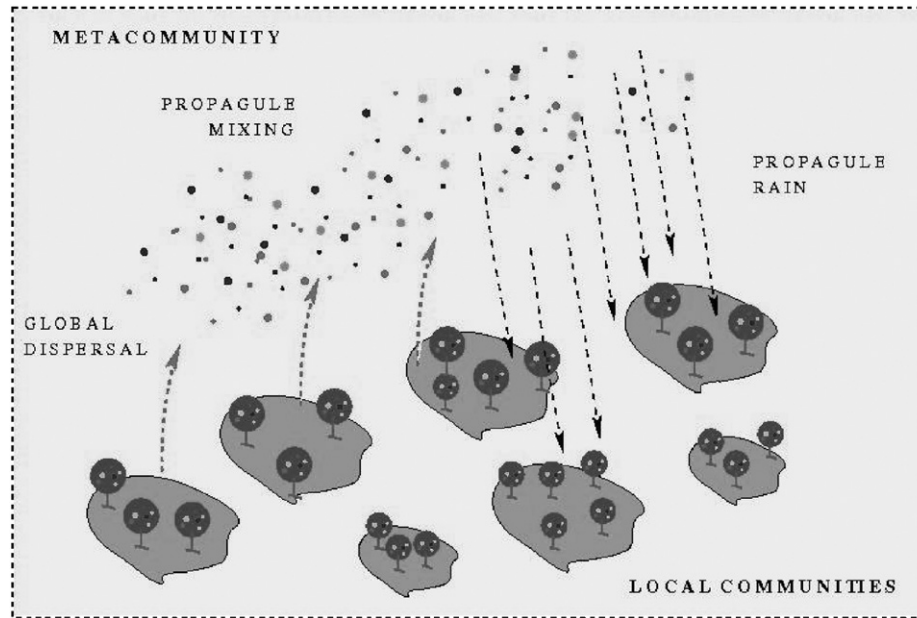


Fig. 1. Classical metacommunity structure. Propagules from all local communities are assumed to completely mix in the metacommunity and return to the local areas in the form of a propagule rain.

J can always be expressed as the following sampling formula:

$$P[n_1, \dots, n_S | J] = \sum_{\{N_1, \dots, N_S\}} \sum_{\{x_1, \dots, x_S\}} P[n_1, \dots, n_S | N_1, \dots, N_S, \Theta, J] \times P[N_1, \dots, N_S | x_1, \dots, x_S] P[x_1, \dots, x_S]. \quad (1)$$

Here, $P[x_1, \dots, x_S]$ is the multivariate distribution of abundances x_1, \dots, x_S in the biogeographical area (or metacommunity, *sensu* Hubbell, 2001) that results from some eco-evolutionary process; $P[N_1, \dots, N_S | x_1, \dots, x_S]$ is the multivariate distribution of abundances N_1, \dots, N_S in the local community that depends on local processes and on the abundances x_1, \dots, x_S in the metacommunity via dispersal; $P[n_1, \dots, n_S | N_1, \dots, N_S, J]$ is the multivariate distribution of abundances n_1, \dots, n_S in the sample of size J that results from the sampling process. Eq. (1) is the central expression of our theory, as it contains all the information on evolutionary, ecological and sampling processes. We call it a general sampling formula for species abundances (or allelic frequencies). By performing the sum over the metacommunity abundances Eq. (1) can also be written as

$$P[n_1, \dots, n_S | J] = \sum_{\{N_1, \dots, N_S\}} P[n_1, \dots, n_S | N_1, \dots, N_S, J] \times P[N_1, \dots, N_S]. \quad (2)$$

This expression accounts for sampling from a given community described by $P[N_1, \dots, N_S]$. One-dimensional counterparts of this expression have been presented in previous sampling theories (Pielou, 1969; Dewdney, 1998; Lande et al., 2003). Alternatively we can perform the sum

over local abundances N_1, \dots, N_S and write

$$P[n_1, \dots, n_S | J] = \sum_{\{x_1, \dots, x_S\}} P[n_1, \dots, n_S | x_1, \dots, x_S, J] \times P[x_1, \dots, x_S]. \quad (3)$$

This expression relates the local sample directly to the metacommunity under different sampling assumptions (Alonso and McKane, 2004; Etienne and Alonso, 2005).

In community ecology, the importance of multivariate abundance distributions has been recognized only recently (Etienne, 2005; Etienne and Alonso, 2005; Etienne and Olf, 2004, 2005; Alonso et al., 2006; Chave et al., 2006). Multivariate abundance distributions contain all measurable quantities that we can extract from local samples, such as the expected number of species with a given abundance in the sample (Etienne and Alonso, 2005) which is the classical species abundance distribution, species covariation across samples, community similarity between samples, etc. Sampling formulas are exact likelihoods and can thus be used in parameter estimation (by maximum likelihood or Bayesian approaches, Etienne and Olf, 2005; Etienne et al., 2006) and model comparison (Etienne and Olf, 2005; Chave et al., 2006). In neutral theory, examples of sampling formulas are the Ewens (1972) and Etienne (2005, 2007) sampling formulas. For instance, the Etienne sampling formula takes the form of (3), where $P[x_1, \dots, x_S]$ describes the extinction-speciation stationary distribution of relative species abundances in a neutral metacommunity (Ewens, 1972) and $P[n_1, \dots, n_S | x_1, \dots, x_S]$ is a dispersal-limited sampling distribution (Etienne and Alonso, 2005).

Table 1
Explanation of mathematical symbols

| Symbol | Explanation |
|---|---|
| n_k, N_k, x_k | Abundance of species k in the sample, the local community and the metacommunity |
| p_k, \hat{p}_k, \vec{p} | Relative abundance and rescaled relative abundance of species k in a species pool, and the vector of relative abundances (this is an abstract pool for the metacommunity model, and it is the metacommunity for the local community model) |
| J, J_L, J_M | Sum of abundances in the sample, the local community and the metacommunity (i.e. their respective sizes) |
| $J_{L,k}, J_{M,k}$ | Sum of abundance of all species in the local community and metacommunity, except species $1 \dots k - 1$, Eq. (12) |
| S | Number of species in the sample |
| S_T | Total number of species that can possibly exist (may be infinite) |
| Φ_j | The number of species in the sample that have abundance j |
| $P[]$ | Probability, sampling formula |
| $P_L[], P_U[]$ | Labeled and unlabeled version of the sampling formula, Eqs. (8) and (9) |
| $t, \hat{t}, T_{zero-sum}, T_{non-zero-sum}, T$ | Time, rescaled (dimensionless) time, characteristic time scale for zero-sum dynamics and non-zero-sum dynamics, number of censuses |
| $g_i, r_i, \hat{g}_i, \hat{r}_i$ | Probability rates of increase and decrease in abundance and their rescaled versions |
| Θ | A general symbol for model parameters |
| $\sigma, \sigma_c, \lambda$ | Speciation rate, per capita speciation rate and immigration rate (probability of speciation/immigration per unit time) |
| θ, I | Fundamental biodiversity number and fundamental dispersal number, Eqs. (15), (B.5), and (22) |
| v, m | Dimensionless speciation and immigration rates (probabilities), Eqs. (B.4) and (B.15) |
| \hat{I}_k | Fundamental dispersal number rescaled by relative abundances of species $1 \dots k - 1$ in the metacommunity, Eq. (B.20) |
| β, δ | Birth rate and death rate (probability of birth/death per unit time) |
| $R, R(t)$ | Reproduction number and reproduction number as a function of time, Eqs. (16) and (28) |
| $\pi(\dots)$ | Permutation |
| $\Omega(p_k), f(p_k)$ | Density and probability density of species k between relative abundance p_k and $p_k + dp_k$, Eqs. (33) and (34) |
| K | Rescaling parameter, Eqs. (B.3) and (B.13) |

To obtain expressions for $P[N_1, \dots, N_S | x_1, \dots, x_S]$ and $P[x_1, \dots, x_S]$ we make use of the master equation approach (Nisbet and Gurney, 1982; Renshaw, 1991; McKane et al., 2000). This approach has been extensively and successfully used in the context of neutral community ecology (Volkov et al., 2003; Vallade and Houchmandzadeh, 2003; McKane et al., 2004), but it is quite uncommon in population genetics where diffusion approaches are more popular (Wright, 1969). A general stochastic model for the

abundance n of a species can be written as

$$\frac{dP[n, t]}{dt} = g_{n-1}P[n-1, t] + r_{n+1}P[n+1, t] - (g_n + r_n)P[n, t], \quad (4)$$

where $P[n, t]$ is the probability that a species has abundance n at time t , and g_n and r_n are the rates at which the species increases or decreases in abundance (due to birth, death, immigration, emigration or speciation). We will denote the equilibrium distribution by $P[n]$ (i.e. simply dropping the argument t) and it is given by (Goel and Richter-Dyn, 1974; Van Kampen, 1992):

$$P[n] = P[0] \frac{g_0}{r_n} \prod_{i=1}^{n-1} \frac{g_i}{r_i}, \quad n > 0. \quad (5)$$

We can calculate $P[0]$ using the requirement that $\sum_{n \geq 0} P[n] = 1$:

$$P[0] = \left(1 + \sum_{n > 0} \frac{g_0}{r_n} \prod_{i=1}^{n-1} \frac{g_i}{r_i} \right)^{-1}. \quad (6)$$

The master equation thus gives us an expression for $P[n]$ where n is the abundance in the metacommunity or, when local dynamics are modeled, for $P[n|x]$ where n is the abundance in the local community and x is the abundance in the metacommunity. From this we may build expressions for the multivariate probabilities $P[x_1, \dots, x_S]$ and $P[N_1, \dots, N_S | x_1, \dots, x_S]$. For our sampling theory, however, we are interested in obtaining the probability distribution $P[n_1, \dots, n_S | J]$ for a sample of size J , rather than for the full community, and (1) tells us that we need an expression for $P[n_1, \dots, n_S | N_1, \dots, N_S, J]$. This of course depends on the sampling process. Although other sampling procedures can be implemented, in most realistic cases we sample randomly without replacement. Then $P[n_1, \dots, n_S | N_1, \dots, N_S, J]$ is given by the multivariate hypergeometric distribution (Etienne and Alonso, 2005):

$$P[n_1, \dots, n_S | N_1, \dots, N_S, J] = \frac{\prod_{k=1}^S \binom{N_k}{n_k}}{\binom{J_L}{J}}, \quad (7)$$

where J_L is the local community size, $J_L = \sum_{k=1}^S N_k$.

If we are only interested in the distribution of species abundances, and not so much in the abundances of particular species, we say that species are unlabeled (Tavaré and Ewens, 1997; Etienne and Alonso, 2005), otherwise we say that they are labeled. Let us assume that there are S_T species that we can possibly observe; S_T is typically infinite, as we shall see below. We can write the following relationship between P_U , the probability of observing an unlabeled species abundance distribution,

and P_L , the probability of observing abundances of specified species:

$$P_U[n_1, \dots, n_S | J] = \sum_{\{\pi(n_1, \dots, n_{S_T})\}} P_L[n_1, \dots, n_{S_T} | J], \quad (8)$$

where the sum is over all possible permutations of the observed abundances n_1, \dots, n_S over the species $1, \dots, S_T$ (the remaining $S_T - S$ species receive abundance 0). When species are exchangeable (symmetric), $P_L[n_1, \dots, n_{S_T} | J]$ is the same for all permutations and hence (8) simplifies to

$$P_U[n_1, \dots, n_S | J] = \frac{S_T!}{\Phi_0! \prod_{j=1}^J \Phi_j!} P_L[n_1, \dots, n_S, 0, 0, 0, \dots | J], \quad (9)$$

where Φ_j is the number of species that have abundance j . Again, unobserved species naturally have abundance 0. For convenience, we have numbered the species on the right-hand side so that we first list the abundances of the observed species and then the unobserved species, but this is arbitrary, as long as it is understood that interchanging two unequal abundances generates a different distribution of labeled species, but the same distribution of unlabeled species. We will drop the subscripts U and L for notational convenience. Whenever zeros are added, it is clear that the labeled version is meant, otherwise we are dealing with the unlabeled version.

We noted above that we need to build expressions for the multivariate probabilities $P[N_1, \dots, N_S | x_1, \dots, x_S]$ and $P[x_1, \dots, x_S]$ from the results of the master equation approach. These results of course depend on the model. We will consider two general models. The first is a community consisting of species with completely independent dynamics: the species do not “feel” the presence of heterospecifics. The second is a community with a constant total number of individuals: a decrease of the abundance of one species must be followed by an increase in the same or another species. This is a community subject to the zero-sum constraint. Because we are considering models we will also add to our expressions a parameter Θ to denote the model parameters.

2.1. Independent species

When species in a community grow independently of one another, $P[x_1, \dots, x_S]$ and $P[N_1, \dots, N_S | x_1, \dots, x_S]$ are given by

$$P[x_1, \dots, x_S, 0, 0, 0, \dots | \Theta] = \prod_{k=1}^S P[x_k | \Theta] \prod_{k=S+1}^{S_T} P[0 | \Theta], \quad (10a)$$

$$P[N_1, \dots, N_S, 0, 0, 0, \dots | \Theta] = \prod_{k=1}^S P[N_k | \Theta, x_k] \prod_{k=S+1}^{S_T} P[0 | \Theta, x_k]. \quad (10b)$$

2.2. Dependent species: the subsample approach

When species in a community experience the population sizes of other species and of themselves, the stochastic variables n_k are not independent and (10) does not apply. We may then use the subsample approach (Etienne and Alonso, 2005):

$$P[x_1, \dots, x_S, 0, 0, 0, \dots | \Theta, J_M] = \prod_{k=1}^{S_T} P[x_k | \Theta, x_1, \dots, x_{k-1}] = \prod_{k=1}^{S_T} P[x_k | \hat{\Theta}_k, J_{M,k}], \quad (11a)$$

$$P[N_1, \dots, N_S, 0, 0, 0, \dots | \Theta, J_L] = \prod_{k=1}^{S_T} P[N_k | \Theta, x_k, N_1, \dots, N_{k-1}] = \prod_{k=1}^{S_T} P[N_k | \hat{\Theta}_k, x_k, J_{L,k}], \quad (11b)$$

where

$$J_{M,k} := J_M - \sum_{i=1}^{k-1} x_i, \quad (12a)$$

$$J_{L,k} := J_L - \sum_{i=1}^{k-1} N_i \quad (12b)$$

is the (meta- or local) community size that is still available to species k , and where $\hat{\Theta}_k$ also incorporates the fact that n_i is known for all species $i = 1, \dots, k-1$. This will become clear below. The subsample approach is nicely illustrated by a very simple example, the extension of the binomial distribution to the multinomial distribution. See Appendix A.

3. Non-zero-sum neutral communities

We will apply our theory to neutral models for the metacommunity (Hubbell, 2001) where there is only speciation, and to the local community, where there is only immigration from the metacommunity, in addition to the standard birth and death processes. We will frequently use the Pochhammer notation

$$(\alpha)_n := \prod_{i=1}^n (\alpha + i - 1) = \prod_{i=0}^{n-1} (\alpha + i). \quad (13)$$

For $n = 0$ we adhere to the convention that $(\alpha)_0 = 1$.

Here we only derive the sampling formulas for the non-zero-sum case, as they are well-known for the zero-sum case (Ewens, 1972; Etienne, 2005). For completeness, however, Appendix B gives the full derivation for the zero-sum case using the subsample approach; this derivation is different from the derivation originally used to obtain the sampling formulas in the zero-sum case and it is instructive to see that, except for the subsample approach, the derivation is analogous to the non-zero-sum case.

3.1. Metacommunity

The distribution of species relative abundances in a neutral metacommunity in speciation–extinction equilibrium was given by Ewens (1972). This distribution is both a sampling formula and a full description of species abundances in a neutral metacommunity. While this distribution was found under the assumption of constant community size, i.e. zero-sum dynamics, here we show that it can be derived without making this assumption, but instead assuming independent species with linear (density independent) growth rates.

In the metacommunity there is only speciation, in addition to birth and death processes. We treat speciation as immigration from a species pool. Treating speciation as an immigration process is only realistic in the limit of an infinite metacommunity. A model that explicitly accounts for finite metacommunity size in speciation is much more complicated, unless metacommunity size is held constant, which is the zero-sum case of Appendix B. The one-step stochastic birth–death–immigration model formulated in Eqs. (14) and (21), was first fully solved by Kendall (1948). He introduced this birth–death–immigration process to give a mechanistic biological basis to Fisher’s empirical logseries distribution when immigration is very low (McKane et al., 2000). More recently, the same argument has been used in the context of neutral theory to give an evolutionary explanation to Fisher’s logseries in the metacommunity through neutral speciation (Volkov et al., 2003). Here we give a more complete description and show that it can also be used to derive the corresponding sampling formula. The derivation of this sampling formula by taking the appropriate limit $S_T \rightarrow \infty$ was suggested by Hoppe (1987) and Costantini and Garibaldi (2004), but here it is made explicit in the immigration model for the first time.

We assume that there are S_T species in the species pool with $S_T \rightarrow \infty$ and each species in the pool having relative abundance $p_k = p \rightarrow 0$, such that $\sum_{k=1}^{S_T} p_k = S_T p = 1$ (hence $p = 1/S_T$). We denote the speciation rate by σ and birth and death rates by β and δ , respectively. Because we are dealing with neutral species, these rates (and also the immigration rate λ below) are assumed to be identical for all species. We will show, however, that this assumption can be somewhat relaxed without changing the results. A species grows in abundance if a birth event happens in its population or when it speciates from the species pool. Because $p \rightarrow 0$, the probability that a species already present in the metacommunity enters the metacommunity through speciation approaches zero, but for new species (of which there are infinitely many: $S_T - S$) this probability is non-zero. As stated above, we assume here that species grow independently. Then, the dynamics of each species can be described by (4) with the following rates of increase and decrease:

$$g_x = \beta x + \sigma p, \tag{14a}$$

$$r_x = \delta x. \tag{14b}$$

We now define

$$\theta := \frac{\sigma}{\beta}. \tag{15}$$

In the limit of an infinite metacommunity, as we described above, this parameter is the non-zero-sum expression of the well-known fundamental biodiversity number (Hubbell, 2001). In this limit, it has the same biological interpretation, namely as the number of new species arising in the system per birth. We further define

$$R := \frac{\beta}{\delta} \tag{16}$$

which has the interpretation of a net reproduction number, i.e. the number of births per death. The solution for $P[x_k]$, $x_k > 0$, is, using (5) and (6),

$$\begin{aligned} P[x_k] &= P_k[0] \frac{g_0}{r_{x_k}} \prod_{i=1}^{x_k-1} \frac{g_i}{r_i} \\ &= (1 - R)^{\theta p} \frac{R^{x_k}}{x_k} \frac{(\theta p)_{x_k}}{(x_k - 1)!}, \end{aligned} \tag{17}$$

where $P_k[0] = (1 - R)^{\theta p}$.

Because we have independent species, we can apply (10) to obtain

$$\begin{aligned} P[x_1, \dots, x_S, 0, 0, 0, \dots | \theta] &= \left(\prod_{k=1}^S R^{x_k} \frac{(\theta p)_{x_k}}{x_k!} (1 - R)^{\theta p} \right) \\ &\quad \times \left(\prod_{k=S+1}^{S_T} (1 - R)^{\theta p} \right). \end{aligned} \tag{18}$$

For a random sample from the metacommunity the sampling formula is given by (3) where $P[n_1, \dots, n_S | x_1, \dots, x_S]$ is the hypergeometric distribution (7).

$$\begin{aligned} P[n_1, \dots, n_S, 0, 0, 0, \dots | \theta, J] &= \sum_{x_1, \dots, x_S} \frac{\prod_{k=1}^S \binom{x_k}{n_k} \frac{(\theta p)_{n_k}}{n_k!} (1 - R)^{\theta p} R^{x_k} (\prod_{k=S+1}^{S_T} (1 - R)^{\theta p})}{\binom{J_M}{J}} \\ &= J! \sum_{x_1, \dots, x_S} \left(\prod_{k=1}^{S_T} (1 - R)^{\theta p} \right) \frac{(J_M - J)!}{J_M!} \prod_{k=1}^S \frac{(\theta p)_{x_k}}{n_k! (x_k - n_k)!} R^{x_k} \\ &= J! \sum_{x_1, \dots, x_S} (1 - R)^{\theta} \frac{(J_M - J)!}{J_M!} R^{J_M} \prod_{k=1}^S \frac{(\theta p)_{n_k} (\theta p + n_k)_{x_k - n_k}}{n_k! (x_k - n_k)!} \\ &= J! \prod_{k=1}^S \frac{(\theta p)_{n_k}}{n_k!} \sum_{J_M} \frac{(1 - R)^{\theta} R^{J_M}}{J_M!} \\ &\quad \times \left[(J_M - J)! \sum_{(x_1, \dots, x_S) \sum_{k=1}^S x_k = J_M} \prod_{k=1}^S \frac{(\theta p + n_k)_{x_k - n_k}}{(x_k - n_k)!} \right] \end{aligned}$$

$$\begin{aligned}
 &= J! \prod_{k=1}^S \frac{(\theta p)_{n_k}}{n_k!} \left(\sum_{J_M} \frac{(1-R)^\theta R^{J_M}}{J_M!} \right) \\
 &\quad \times \left[\sum_{(K_1, \dots, K_S) \sum_{k=1}^S K_k = K} K! \prod_{k=1}^S \frac{(\theta p + n_k)_{K_k}}{K_k!} \right] \\
 &= \frac{J!}{(\theta)_J} \prod_{k=1}^S \frac{(\theta p)_{n_k}}{n_k!} \left(\sum_{J_M} \frac{(1-R)^\theta R^{J_M}}{J_M!} \right) (\theta)_J [(\theta + J)_{J_M - J}] \\
 &= \frac{J!}{(\theta)_J} \prod_{k=1}^S \frac{(\theta p)_{n_k}}{n_k!} \left(\sum_{J_M} \frac{(\theta)_{J_M} (1-R)^\theta R^{J_M}}{J_M!} \right) \\
 &= \frac{J!}{(\theta)_J} \prod_{k=1}^S \frac{(\theta p)_{n_k}}{n_k!}. \tag{19}
 \end{aligned}$$

Using (9) and noting that $\Phi_0 = S_T - S$ we obtain

$$\begin{aligned}
 &P[n_1, \dots, n_S | \theta, J] \\
 &= \frac{S_T!}{\Phi_0! \prod_{j=1}^{J_M} \Phi_j!} P[n_1, \dots, n_S, 0, 0, 0, \dots | \theta, J] \\
 &= \frac{S_T!}{\Phi_0! \prod_{j=1}^{J_M} \Phi_j!} J! \prod_{k=1}^S \frac{(\theta p)_{n_k}}{n_k!} \frac{1}{(\theta)_J} \\
 &= \frac{J! S_T!}{(S_T - S)! \prod_{j=1}^J \Phi_j!} \frac{(\theta p)^S}{(\theta)_J} \left(\prod_{k=1}^S \frac{1}{n_k} \frac{(\theta p + 1)_{n_k - 1}}{(n_k - 1)!} \right) \\
 &\rightarrow \frac{J!}{\prod_{k=1}^S n_k \prod_{j=1}^J \Phi_j!} \frac{\theta^S}{(\theta)_J} \tag{20}
 \end{aligned}$$

because in the limit that we are considering ($S_T \rightarrow \infty, p \rightarrow 0, S_T p \rightarrow 1$) we have $(S_T! / (S_T - S)!) p^S \rightarrow (S_T p)^S \rightarrow 1$ and $(\theta p + 1)_{n_k - 1} \rightarrow (1)_{n_k - 1} = (n_k - 1)!$. In (20) we recognize the [Ewens \(1972\)](#) sampling formula. Note that the equality $(\theta p)_{n_k} = \theta p (\theta p + 1)_{n_k - 1}$ that is used in this equation is only valid for $n_k > 0$ which is indeed true for n_1, \dots, n_S (if any one of these were equal to 0, the list would be $n_1, \dots, n_{S-1}, 0, 0, 0, \dots$).

Note that our derivation does not require that all species have the same speciation, birth and death rates; we only require that they have the same R and θ . The same θ means that the speciation rate should be proportional to the birth rate, see (15).

3.2. Local community

We consider a local community with immigration from an infinite metacommunity containing S_T species that have relative abundances p_k . There is no local speciation. We assume again that species grow independently with the following rates of increase and decrease:

$$g_N = \beta N + \lambda p_k, \tag{21a}$$

$$r_N = \delta N, \tag{21b}$$

where λ is the immigration rate, β is the birth rate and δ is the death rate. We now define

$$I := \frac{\lambda}{\beta}. \tag{22}$$

This number is the non-zero-sum fundamental dispersal number and has the interpretation of number of immigrant individuals per local birth. Eqs. (21) become

$$g_N = \beta(N + I p_k), \tag{23a}$$

$$r_N = \delta N. \tag{23b}$$

As in the metacommunity non-zero-sum case, we introduce the net reproduction number (16). Using (5) and (6), one finds the following solution for $P[N_k]$:

$$\begin{aligned}
 P[N_k] &= P_k[0] \frac{I p_k}{\delta n_k} \prod_{i=1}^{N_k - 1} \frac{\beta(i + I p_k)}{\delta i} \\
 &= (1 - R)^{I p_k} R^{N_k} \frac{(I p_k)_{N_k}}{n_k!}. \tag{24}
 \end{aligned}$$

Because we have independent species, we can apply (10):

$$\begin{aligned}
 &P[N_1, \dots, N_S, 0, 0, 0, \dots | I, \vec{p}] \\
 &= (1 - R)^{I \sum p_k} \prod_{k=1}^S \frac{(I p_k)_{N_k}}{N_k!} R^{N_k}. \tag{25}
 \end{aligned}$$

Note that in this expression species must be labeled because their metacommunity abundances are assumed to be given by \vec{p} . As we are sampling from a local community where these metacommunity abundances are given, we can apply (2) with the hypergeometric distribution (7) for $P[n_1, \dots, n_S | N_1, \dots, N_S]$. The derivation is completely analogous to that of (19); we only need to replace θ by I , p by p_k , J_M by J_L and x_k by N_k :

$$\begin{aligned}
 &P[n_1, \dots, n_S, 0, 0, 0, \dots | I, J, \vec{p}] \\
 &= \frac{J!}{(I)_J} \prod_{k=1}^S \frac{(I p_k)_{n_k}}{n_k!}. \tag{26}
 \end{aligned}$$

The probability distribution in Eqs. (B.22) and (26) are known as the compound multinomial Dirichlet distribution or the multivariate negative hypergeometric distribution ([Mosimann, 1962](#); [Rannala, 1996](#)). In the context of neutral biodiversity theory we propose to call this distribution the “dispersal-limited multinomial (DLM)” because of the interpretation of its parameter I as the fundamental dispersal number ([Etienne and Alonso, 2005, 2006](#)).

Eq. (26) is actually also valid when equilibrium has not been reached. Model (21) with initial condition $P[N = 0 | t = 0] = 1$ has the following full solution:

$$P[N_k, t] = (1 - R(t))^{I p_k} R(t)^{N_k} \frac{(I p_k)_{N_k}}{N_k!}, \tag{27}$$

where

$$R(t) = \frac{\beta(e^{(\beta-\delta)t} - 1)}{\beta - \delta} = R \frac{1 - e^{(\beta-\delta)t}}{1 - Re^{(\beta-\delta)t}}, \quad (28)$$

where R is given by (16). Eq. (27) has exactly the same form as (24) with $R(t)$ replacing R , and because R disappears when conditioning on J_L , we again obtain (26). Note that when $\beta < \delta$ then for $t \rightarrow \infty$, $R(t) \rightarrow R$; when $\beta > \delta$ there is no equilibrium, as the populations of all species grow exponentially. Nevertheless, on the way to infinite population sizes, sampling formula (26) still applies.

Although some effort has been made to approximate the time-dependent behavior of the zero-sum local model (McKane et al., 2004), an exact solution is not yet available. However, we can compare the dynamics of the zero-sum model and the non-zero-sum model by solving the differential equation for the *expected* population size of a species k , $\langle N_k \rangle$,

$$\frac{d\langle N_k \rangle}{dt} = \langle g_{N_k} \rangle - \langle r_{N_k} \rangle. \quad (29)$$

When the population size is initially zero (the species will immigrate sooner or later), then we have

$$\langle N_k(t) \rangle_{zero-sum} = J_L p_k (1 - e^{-(\lambda/J_L)t}), \quad (30a)$$

$$\langle N_k(t) \rangle_{non-zero-sum} = \frac{\lambda p_k}{\delta - \beta} (1 - e^{-(\delta-\beta)t}). \quad (30b)$$

We can define a characteristic time T for both colonization processes:

$$T_{zero-sum} := \frac{J_L}{\lambda}, \quad (31a)$$

$$T_{non-zero-sum} := \frac{1}{\delta - \beta}. \quad (31b)$$

Not surprisingly, these times are different; the characteristic time scale of a species in the zero-sum community is set by the immigration rate, whereas the time scale of a species in the non-zero-sum community is governed by the birth and death rates.

3.3. Linking the local community to the metacommunity

So far we have derived a sampling distribution where a set of relative abundances in the metacommunity, described by the relative abundance vector (p_1, \dots, p_S) , is assumed to be fixed and constant. Our formulae then consider labeled species, as the species in the local community (sample) must correspond to the species in the metacommunity. Usually, we do not know the relative abundances in the metacommunity, \vec{p} , but we often have a model for the metacommunity that provides the probability density function of \vec{p} . Let us call this $f(\vec{p})$. We can then write an unlabeled version by integrating over all

possible vectors \vec{p} ,

$$P[n_1, \dots, n_S | I, J] = \frac{S_T!}{\Phi_0! \prod_{j=1}^J \Phi_j!} \int_0^1 \int_0^{1-p_1} \dots \times \int_0^{1-\sum_{k=1}^{S_T-1} p_k} \frac{J!}{(I)_J} \prod_{k=1}^{S_T} \frac{(I p_k)_{n_k}}{n_k!} f(\vec{p}) d\vec{p}. \quad (32)$$

For a neutral metacommunity the number of species with relative abundance between p_k and $p_k + dp_k$ is given by (in the limit that the total number of species S_T goes to infinity, Ewens, 1972)

$$\Omega(p_k) dp_k = \frac{\theta(1-p_k)^{\theta-1}}{p_k} dp_k. \quad (33)$$

Because a probability distribution is the same as the corresponding frequency distribution of an infinite population, this implies that the probability that a particular species, selected from the metacommunity (of infinite size), has relative abundance between p_k and $p_k + dp_k$ is given by (using (A.3) to define \hat{p}_1)

$$f(p_1) dp_1 = \frac{\theta(1-p_1)^{\theta-1}}{p_1} dp_1 = f(\hat{p}_1) d\hat{p}_1. \quad (34)$$

For a second species, because of neutrality this probability has the same form when conditioned on the relative abundance of the first, p_1 :

$$f(p_2 | p_1) dp_2 = \frac{\theta(1-\frac{p_2}{1-p_1})^{\theta-1}}{\frac{p_2}{1-p_1}} \frac{1}{1-p_1} dp_2 = \frac{\theta(1-\frac{p_2}{1-p_1})^{\theta-1}}{\frac{p_2}{1-p_1}} \frac{1}{1-p_1} dp_2 = \frac{\theta(1-\hat{p}_2)^{\theta-1}}{\hat{p}_2} d\hat{p}_2 = f(\hat{p}_2 | \hat{p}_1) d\hat{p}_2, \quad (35)$$

where \hat{p}_2 is defined by (A.3). Likewise, for S_T species selected from the metacommunity, we have

$$f(p_1, \dots, p_{S_T}) dp_1 \dots dp_{S_T} = f(\hat{p}_1, \dots, \hat{p}_{S_T}) d\hat{p}_1 \dots d\hat{p}_{S_T} = \prod_{k=1}^{S_T} \frac{\theta(1-\hat{p}_k)^{\theta-1}}{\hat{p}_k} d\hat{p}_k = \frac{1}{\left(\int_0^1 \frac{\theta(1-\hat{p})^{\theta-1}}{\hat{p}} d\hat{p} \right)^{S_T}} \prod_{k=1}^{S_T} \frac{\theta(1-\hat{p}_k)^{\theta-1}}{\hat{p}_k} d\hat{p}_k. \quad (36)$$

Inserting (36) into (32) yields

$$\begin{aligned}
& P[n_1, \dots, n_S | I, \theta, J] \\
&= \frac{S_T!}{(S_T - S)! \prod_{j=1}^J \Phi_j!} \frac{1}{\left(\int_0^1 \frac{\theta(1-\hat{p})^{\theta-1}}{\hat{p}} d\hat{p} \right)^{S_T}} \\
&\quad \times \int_0^1 \dots \int_0^1 \frac{J!}{(I)_J} \prod_{k=1}^S \left(\frac{(\hat{I}_k \hat{p}_k)_{n_k}}{n_k!} \frac{\theta(1-\hat{p}_k)^{\theta-1}}{\hat{p}_k} d\hat{p}_k \right) \\
&\quad \times \prod_{k=S+1}^{S_T} \left(\frac{(\hat{I}_k \hat{p}_k)_0}{0!} \frac{\theta(1-\hat{p}_k)^{\theta-1}}{\hat{p}_k} d\hat{p}_k \right) \\
&= \frac{J!}{\prod_{k=1}^S n_k! \prod_{j=1}^J \Phi_j!} \frac{\theta^S}{(I)_J} \int_0^1 \dots \int_0^1 \\
&\quad \times \prod_{k=1}^S \left(\frac{(\hat{I}_k \hat{p}_k)_{n_k}}{n_k!} \frac{(1-\hat{p}_k)^{\theta-1}}{\hat{p}_k} d\hat{p}_k \right) \\
&\quad \times \frac{\prod_{k=S+1}^{S_T} \left(\int_0^1 \frac{\theta(1-\hat{p}_k)^{\theta-1}}{\hat{p}_k} d\hat{p}_k \right)}{\left(\int_0^1 \frac{\theta(1-\hat{p})^{\theta-1}}{\hat{p}} d\hat{p} \right)^{S_T-S}} \\
&\quad \times \frac{S_T!}{(S_T - S)! \left(\int_0^1 \frac{\theta(1-\hat{p})^{\theta-1}}{\hat{p}} d\hat{p} \right)^S} \\
&\rightarrow \frac{J!}{\prod_{k=1}^S n_k! \prod_{j=1}^J \Phi_j!} \frac{\theta^S}{(I)_J} \int_0^1 \dots \int_0^1 \\
&\quad \times \prod_{k=1}^S \left(\frac{(\hat{I}_k \hat{p}_k)_{n_k}}{n_k!} \frac{(1-\hat{p}_k)^{\theta-1}}{\hat{p}_k} d\hat{p}_k \right), \tag{37}
\end{aligned}$$

where we have used the fact that

$$\int_0^1 \frac{\theta(1-\hat{p})^{\theta-1}}{\hat{p}} d\hat{p} = S_T \quad \text{and that} \quad \frac{S_T!}{(S_T - S)!} \rightarrow S_T^S$$

as S_T approaches infinity. Eq. (37) is identical to the Etienne sampling formula (Etienne and Alonso, 2005). Because the probability density $f(\vec{p})$ is the same for both the zero-sum and the non-zero-sum model, the Etienne sampling formula also applies to the non-zero-sum case. By using relative abundances for the metacommunity, we have implicitly assumed the size of the metacommunity to be infinite. However, the Etienne sampling formula, (37), also applies when the metacommunity is finite (Etienne and Alonso, 2005).

In sum, the Etienne sampling formula (37) gives the probability of obtaining a data set from a dispersal-limited local neutral community regardless of the zero-sum constraint: the same formula applies, on the one hand, to zero-sum local communities in stochastic equilibrium, and, on the other hand, to non-zero-sum neutral communities with independent species having density-independent per capita birth–death rates, at any point in time, that is, even when the local community has not yet reached stochastic

equilibrium. We remark that, formally, we have considered only one form of non-zero-sum dynamics, namely where species are completely independent, which is a simple and natural choice, but there are other forms of non-zero-sum dynamics where species are dependent to some extent but not completely restricted by a fixed community size. It remains to be shown what the sampling formulas are in those cases.

4. Discussion

Since Hubbell published his monograph in 2001, the evaluation of neutral theory has encountered three main difficulties. First, neutral theory in community ecology has been formulated using a variety of models (Chave, 2004; Alonso et al., 2006). Second, most studies based on abundance data have focused on the expected abundance curve, better known as the species abundance distribution (McGill, 2003; Volkov et al., 2003; Turnbull et al., 2005; Volkov et al., 2005; Harpole and Tilman, 2006). Although this curve is a signature of community structure (Pueyo, 2006), recent work has emphasized the difficulty of identifying the underlying processes leading to this pattern if a snapshot of the abundance curve is the only information available (Chave and Leigh, 2002; Bell, 2005; Purves and Pacala, 2005; Volkov et al., 2005; Etienne et al., 2006). Third, alternative niche theories were only loosely defined (McGill, 2003); they did not generally provide the probability of obtaining data through sampling assuming the alternative niche model at play. As a consequence, it has been practically impossible to perform sound model selection (Chave et al., 2006). The sampling theory presented in this paper aims to find a solution to these three difficulties. We discuss them in order.

First, we have stressed that Hubbell's basic model (Hubbell, 2001) plays a central role in the theory. This model was initially understood as a mainland–island model (MacArthur and Wilson, 1967; Hubbell, 2001) with zero-sum dynamics. The metacommunity was simply regarded as a regional species pool (Etienne, 2005) with the local community either separate from (real mainland–island) or embedded in the metacommunity (continuous landscape), as pointed out by Alonso et al. (2006). Here we have carefully examined the assumptions of this model in two ways. (i) Most importantly, we have demonstrated that all analytical results on the species abundance distribution apply regardless of zero-sum dynamics. This was already suggested by Caswell's simulations (Caswell, 1976) simulations, argued analytically by Rannala (1996) for the birth–death–immigration model in a population genetics context and hinted at by Volkov et al. (2003, 2005) for the expected abundance curve, but we have provided a more rigorous proof for the full sampling formula (that involves both immigration to the local community and speciation in the metacommunity) using a general sampling theory. Because natural communities are not constrained by a strict zero-sum rule, this result increases the degree of

realism and makes the theory more robust and appealing. (ii) En passant, we have also clarified a different interpretation of this model within the context of an emerging paradigm, the metacommunity concept (Leibold et al., 2004; Holyoak et al., 2005). The model has been applied not only to pure mainland–island systems, but also (and more frequently) to metacommunities in the true sense of the word (e.g. Dornelas et al., 2006): local communities that grow on fragmented patchy areas (Alonso and Pascual, 2006), and that are connected to each other through migration, which is *global* in the model. So, partially isolated local communities host different species brought together by random dispersal from a pool that is only governed by evolutionary forces (speciation). Note that the per capita death rate δ is actually death plus emigration. Any surplus of local production feeds the metacommunity (emigration) and returns to the local community as a propagule rain or global immigration (λ). The crucial assumptions made here are: (a) that emigration does not affect the metacommunity composition and a spatially implicit model can capture the essence of the dynamics and (b) that evolutionary forces act on a different time scale than ecological forces. Under these assumptions—which are not always justified, see e.g. Etienne (2007), Hairston et al. (2005)—the collective behavior of an ensemble of species performing these simple dynamics in the local community is adequately described by the dispersal-limited sampling formula (Etienne, 2005; Etienne and Alonso, 2005, 2006), which gives a measure of the average degree of isolation of local communities by way of I , the fundamental dispersal number, and of the metacommunity diversity by way of θ , the fundamental biodiversity number.

Second, although we agree that the abundance curve may not be enough to elucidate underlying processes, our work on species abundances is based instead on what we call a general sampling formula, the central expression of the theory (1). This is a multivariate abundance distribution that may encode much more information than the simple abundance curve, the expected number of species at each abundance level. Different processes can lead to similar, perhaps indistinguishable, average abundance curves (Volkov et al., 2005). Because this curve is theoretically obtained by averaging the full sampling formula (Etienne and Alonso, 2005), it can be seen as a first moment of a multivariate distribution. But the first moment of a distribution does not describe the distribution completely. So, it is possible that two processes lead to the same average abundance curve (Volkov et al., 2005), but the underlying multivariate distributions may be different (Chave et al., 2006). Moreover, because the multivariate distribution keeps track of individual abundances, it can be extended to multiple samples across space or time. To date, most studies on species abundances dealing with neutral theory have only used the abundance curve (McGill et al., 2006; Pueyo, 2006), rather than this powerful multivariate representation of the community; hence conclusions that

species abundance distributions contain little information are premature. The sampling formula under neutrality for a single sample is the Etienne sampling formula (Etienne, 2005). It has recently been extended to multiple samples across space which conveys more information (Etienne, 2007; Munoz et al., 2007). Also, most studies have analyzed only snapshots of the community rather than dynamical data (but see Gilbert et al., 2006), which may contain more information as well. Our framework can be easily extended to deal with this type of data (38). When we sample the same system at T different times (where the sample size may vary), yielding a time series of multivariate abundance observations $D_{t_1}, D_{t_2}, \dots, D_{t_T}$, we can calculate the likelihood of this time series for the model that is assumed to describe the assembly process, by simply multiplying conditional transitional probabilities:

$$P[D_{t_1}, D_{t_2}, \dots, D_{t_T} | \Theta, J] \\ = P[D_{t_T} | D_{t_T-1}, \Theta, J] \dots P[D_{t_2} | D_{t_1}, \Theta, J] P[D_{t_1} | \Theta, J], \quad (38)$$

where $\Theta = \{\beta, \delta, \lambda\}$ represents model parameters. The probability of observing the first abundance data set, $P[D_{t_1} | \Theta, J]$, can be computed by assuming stochastic equilibrium, or by assuming some initial condition in the past. The actual computation of the conditional probabilities in (38) may be quite involved and requires further theoretical development that is beyond the scope of this paper.

Third, here we have used the fundamental expression (1) to derive sampling formulas in the context of neutral theory, but we recall that this is a general expression that can be extended to situations beyond neutrality as well. We have used the hypergeometric distribution to describe the sampling process, but again our general expression allows other forms for this distribution. Our expression does assume that we sample randomly a predetermined number of individuals. In the field, we can often adapt our sampling strategy to meet this requirement. Some work has been done to deal with the case where a predetermined number of species is sampled instead (Etienne and Olf, 2005). However, the case where an area or a transect is sampled poses serious theoretical challenges. We hope that this paper stimulates research on this problem.

Thus, our sampling theory suggests routes to resolving the three important limitations of recent studies on testing neutral theory vs. niche theories by using species abundance data. However, more interestingly, we provide a sampling description of communities, where both evolutionary and ecological determinants of species coexistence and composition can be analyzed. Hubbell (2001) already studied how species origination influences community patterns. When we analyze our general sampling formula (1) or the particular example given in (32), we notice that this expression consist of two essential pieces. First, we have the multivariate distribution at the metacommunity level ($P(x_1, \dots, x_S)$ in (1) or $f(\vec{p})$ in (32)). Species can potentially diverge in their life history characteristics by

creating and adapting to different biological and physical niches. As a result of such processes some equilibrium between species origination and extinction is generally reached, which can be described by a multivariate distribution. For instance, the Ewens distribution describes this distribution under neutrality and the point mutation mode of speciation. We provide a general method to integrate over the Ewens distribution. A proper change of variables simplifies enormously our integration domain in (32) and leads to the integral expression of Etienne sampling formula (Etienne and Alonso, 2005). However, we can only reach this metacommunity distribution by sampling locally, that is, by looking at local communities. So, the second ingredient of our general sampling formula is a sampling distribution that gives the probability of obtaining a given collection of species abundances given that we know exactly the species composition in the metacommunity ($P[n_1, \dots, n_S | x_1, \dots, x_S, J]$ in (3)). This conditional probability is determined by the sampling process and the ecological processes assumed to play a role during the assembly of local communities. For instance, if dispersal limitation is the main factor in the context of ecologically symmetric species, we obtain the DLM (Etienne and Alonso, 2005) given in (26).

Volkov et al. (2005) have also initiated a sampling theory (see their appendix), borrowing ideas from statistical physics on approximations for very large assemblages. In contrast, our approach is exact and thus enables us to prove the exact equality of zero-sum and non-zero-sum cases in the context of neutral theory, but in other instances it may become very complicated in which case the work by Volkov et al. (2005) presents a promising alternative direction deserving further investigation.

In sum, our work considers current neutral theory as a basis to set up a general sampling theory able to go beyond neutrality and still analyze alternative models in terms of the extent to which empirical data statistically support them. As a first step, in our non-zero-sum example, we provide sampling formulas that assume independence of species. Although this is a valid assumption under neutrality and may be a still reasonable, particularly when only one trophic level is considered, more general community patterns may of course emerge from individual species interactions, described by topologically complex networks. Further integration is needed between sampling theories and network ecology to pave the way for a sampling theory for food webs or mutualistic interaction networks (Bascompte et al., 2003). We believe that our sampling approach can be used to generate (null) models that are necessary in assessing the plausibility of alternative network topologies on the basis of samples of complex communities.

Acknowledgments

We thank two anonymous reviewers for helpful comments. D.A. thanks the support of the James S. McDonnell

Foundation through a Centennial Fellowship to Mercedes Pascual.

Appendix A. The subsample approach: from binomial to multinomial

Let us assume that the relative abundance (frequency) of a species in a(n infinite) community is given by p_k . Then the abundance of a species in a sample with replacement of size J is governed by the binomial distribution with parameters J and p_k :

$$P[n_k | p_k] = \binom{J}{n_k} p_k^{n_k} (1 - p_k)^{J - n_k}. \quad (\text{A.1})$$

Now assume that we have S_T species that have frequencies p_k . The p_k must, by definition, obey $\sum_{k=1}^{S_T} p_k = 1$. This constraint makes the species dependent in the sampling process, and therefore requires the subsample approach. Eq. (11) teaches us that

$$P[n_1, \dots, n_{S_T} | p_1, \dots, p_{S_T}, J] = \prod_{k=1}^{S_T} P[n_k | \hat{p}_k, J_k]. \quad (\text{A.2})$$

In this expression J_k is the sample size minus the abundances of species $i = 1, \dots, k - 1$ (see (12)) and \hat{p}_k is the relative abundance of species k when we exclude all species $i = 1, \dots, k - 1$:

$$\hat{p}_k = \frac{p_k}{1 - \sum_{i=1}^{k-1} p_i}. \quad (\text{A.3})$$

This entails

$$\begin{aligned} 1 - \hat{p}_k &= 1 - \frac{p_k}{1 - \sum_{i=1}^{k-1} p_i} = \frac{1 - \sum_{i=1}^{k-1} p_i - p_k}{1 - \sum_{i=1}^{k-1} p_i} \\ &= \frac{1 - \sum_{i=1}^k p_i}{1 - \sum_{i=1}^{k-1} p_i}. \end{aligned} \quad (\text{A.4})$$

With this reparametrization each $P[n_k | \hat{p}_k, J_k]$ in (A.2) follows the binomial distribution with parameters J_k and \hat{p}_k . Hence, (A.2) becomes

$$\begin{aligned} P[n_1, \dots, n_{S_T} | p_1, \dots, p_{S_T}, J] &= \prod_{k=1}^{S_T} \binom{J_k}{n_k} \hat{p}_k^{n_k} (1 - \hat{p}_k)^{J_k - n_k} \\ &= \prod_{k=1}^{S_T} \frac{J_k!}{n_k! (J_k - n_k)!} \left(\frac{p_k}{1 - \sum_{i=1}^{k-1} p_i} \right)^{n_k} \left(1 - \sum_{i=1}^k p_i \right)^{J_k - n_k} \\ &\quad \times \left(\frac{1}{1 - \sum_{i=1}^{k-1} p_i} \right)^{J_k} \left(\frac{1}{1 - \sum_{i=1}^{k-1} p_i} \right)^{-n_k} \\ &= \prod_{k=1}^{S_T} \frac{J_k!}{n_k! J_{k+1}!} \left(\frac{p_k}{1 - \sum_{i=1}^{k-1} p_i} \right)^{n_k} \end{aligned}$$

$$\begin{aligned} & \times \left(1 - \sum_{i=1}^k p_i\right)^{J_{k+1}} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{-J_k} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{n_k} \\ & = \prod_{k=1}^{S_T} \frac{J_k!}{n_k! J_{k+1}!} p_k^{n_k} \frac{(1 - \sum_{i=1}^k p_i)^{J_{k+1}}}{(1 - \sum_{i=1}^{k-1} p_i)^{J_k}} \\ & = \binom{J}{n_1, \dots, n_{S_T}} \prod_{k=1}^{S_T} p_k^{n_k} \end{aligned} \quad (\text{A.5})$$

and in the last expression we recognize the multinomial distribution with parameters J and p_1, \dots, p_{S_T} .

Appendix B. The subsample approach: zero-sum communities

Here we derive the sampling formulas for the zero-sum case following the same approach as for non-zero-sum communities but making use of the subsample approach to account for the zero-sum constraint. Although the resulting sampling formulas have been derived before in a different way (Ewens, 1972, Karlin and McGregor, 1972; Etienne, 2005), it is instructive to see their derivation in our general framework.

B.1. Metacommunity

If the zero-sum constraint is at work, the rate of increase of a ‘‘focal’’ species represented by x individuals, as given by g_x in (23a), will be depressed by a factor representing the fraction of space that is available for the focal species to grow. This fraction is equal to the fraction occupied by all the other species, $1 - x/J_L$. Furthermore, for zero-sum dynamics the net rate of decrease of the ‘‘focal’’ species, r_x , results from an increase in the abundance of any ‘‘non-focal’’ species in the system. When the total number of individuals in the metacommunity is J_M and there are S_T possible species ($S_T \rightarrow \infty$), we have the following rates of increase and decrease:

$$g_x = \frac{J_M - x}{J_M} \left(\beta x + \frac{\sigma}{S_T} \right), \quad (\text{B.1a})$$

$$r_x = \frac{x}{J_M} \left(\beta(J_M - x) + \sigma \left(1 - \frac{1}{S_T} \right) \right), \quad (\text{B.1b})$$

where σ is the speciation rate (which depends on the constant community size, for example $\sigma = \sigma_c J_M$, where σ_c is the per capita speciation rate). The term $1/S_T$ represents the probability of speciation producing exactly an individual of species k . Note the symmetry between g_x and r_x . The increase of the ‘focal’ species is at the expense of the decrease of any ‘non-focal’ species. This line of reasoning gives a general recipe to turn a symmetric model of population growth into its zero-sum counterpart.

Using (15) we can write (B.1) as

$$g_x = \beta \frac{J_M - x}{J_M} \left(x + \frac{\theta}{S_T} \right), \quad (\text{B.2a})$$

$$r_x = \beta \frac{x}{J_M} \left((J_M - x) + \theta \left(1 - \frac{1}{S_T} \right) \right). \quad (\text{B.2b})$$

If we define

$$K := \sigma + \beta(J_M - 1) \quad (\text{B.3})$$

and

$$v := \frac{\sigma}{K} \quad (\text{B.4})$$

which implies, using (15),

$$\theta := \frac{v}{1 - v} (J_M - 1) \quad (\text{B.5})$$

and we rescale time as

$$\hat{t} := Kt \quad (\text{B.6})$$

then we can also write (B.2) as a dimensionless system that is similar to expressions found in the literature (Hubbell, 2001)

$$\hat{g}_x = \frac{J_M - x}{J_M} \left[\left(\frac{x}{J_M - 1} \right) (1 - v) + \frac{v}{S_T} \right], \quad (\text{B.7a})$$

$$\hat{r}_x = \frac{x}{J_M} \left[\left(\frac{J_M - x}{J_M - 1} \right) (1 - v) + v \left(1 - \frac{1}{S_T} \right) \right]. \quad (\text{B.7b})$$

Inserting (B.2) in (5) gives

$$\begin{aligned} P[x_k | J_M] &= P_k[0] \frac{\frac{J_M \theta}{S_T}}{x_k \left(J_M - n_k + \theta \left(1 - \frac{1}{S_T} \right) \right)} \\ &\times \prod_{i=1}^{x_k-1} \frac{[J_M - i] \left(i + \frac{\theta}{S_T} \right)}{i \left[J_M - i + \theta \left(1 - \frac{1}{S_T} \right) \right]} \\ &= P_k[0] \frac{J_M!}{x_k! (J_M - x_k)!} \frac{\left(\frac{\theta}{S_T} \right)_{x_k}}{\left(\theta \left(1 - \frac{1}{S_T} \right) + J_M - x_k \right)_{x_k}} \\ &= \frac{J_M!}{x_k! (J_M - x_k)!} \frac{\left(\frac{\theta}{S_T} \right)_{x_k} \left(\theta \left(1 - \frac{1}{S_T} \right) \right)_{J_M - x_k}}{(\theta)_{J_M}} \end{aligned} \quad (\text{B.8})$$

so

$$\begin{aligned} P[x_k | \theta, J_{M,k}] &= \frac{J_{M,k}!}{x_k! (J_{M,k} - x_k)!} \frac{\left(\frac{\theta}{S_T} \right)_{x_k} \left(\theta \left(1 - \frac{1}{S_T} \right) \right)_{J_{M,k} - x_k}}{(\theta)_{J_{M,k}}}. \end{aligned} \quad (\text{B.9})$$

Using this in (11) and (9) we obtain

$$\begin{aligned}
 & P[x_1, \dots, x_S | \theta, J_M] \\
 &= \frac{S_T!}{(S_T - S)! \prod_{j=1}^{J_M} \Phi_j!} \prod_{k=1}^{S_T} P[x_k | \theta, J_{M,k}] \\
 &= \frac{S_T!}{(S_T - S)! \prod_{j=1}^{J_M} \Phi_j!} \prod_{k=1}^S \frac{\theta^{J_{M,k}!}}{S_T x_k! (J_{M,k} - x_k)!} \\
 &\quad \times \frac{\left(\frac{\theta}{S_T} + 1\right)^{x_{k-1}} \left(\theta \left(1 - \frac{1}{S_T}\right)\right)^{J_{M,k} - x_k}}{(\theta)_{J_M}} \\
 &\rightarrow \frac{1}{\prod_{j=1}^{J_M} \Phi_j!} \prod_{k=1}^S \frac{\theta^{J_{M,k}!}}{x_k J_{M,k+1}! (\theta)_{J_{M,k}}} \\
 &= \frac{J_M!}{(\theta)_{J_M} \prod_{j=1}^{J_M} \Phi_j!} \prod_{k=1}^S \frac{\theta^{x_k}}{x_k!} \\
 &= \frac{J_M!}{\prod_{k=1}^S x_k! \prod_{j=1}^{J_M} \Phi_j!} \frac{\theta^S}{(\theta)_{J_M}}. \tag{B.10}
 \end{aligned}$$

This formula is invariant under sampling, that is, when we apply (3), then the formula as the same shape; only metacommunity size J_M is replaced by sample size J . Thus we have arrived at Ewens (1972) sampling formula. Because both the zero-sum and the non-zero-sum formulations of the neutral metacommunity model yield the Ewens sampling formula, we have shown that they are identical.

B.2. Local community

We now consider the local community. The total community size is at a constant J_L . The rates of increase and decrease are

$$g_N = \frac{J_L - N}{J_L} [\beta N + \lambda p_k], \tag{B.11a}$$

$$r_N = \frac{N}{J_L} [\beta(J_L - N) + \lambda(1 - p_k)]. \tag{B.11b}$$

Using (22) we can rewrite (B.11) as

$$g_N = \beta \frac{J_L - N}{J_L} [N + I p_k], \tag{B.12a}$$

$$r_N = \beta \frac{N}{J_L} [J_L - N + I(1 - p_k)]. \tag{B.12b}$$

The quantity I is again the fundamental dispersal number (Etienne and Alonso, 2005), to be interpreted as the number of immigrant individuals per local birth. We can make system (B.12) dimensionless by using the scaling parameter K , defined as

$$K := \lambda + \beta(J_L - 1). \tag{B.13}$$

Time is then rescaled to a dimensionless time \hat{t}

$$\hat{t} := Kt \tag{B.14}$$

and the immigration rate λ is rescaled to a dimensionless immigration rate m ,

$$m := \frac{\lambda}{K} = \frac{I}{I + J_L - 1}. \tag{B.15}$$

Inserting this in (B.12), we obtain the set of found in the literature (Vallade and Houchmandzadeh, 2003; McKane et al., 2004),

$$\hat{g}_N = \frac{g_N}{K} = \frac{J_L - N}{J_L} \left[(1 - m) \left(\frac{N}{J_L - 1} \right) + m p_k \right], \tag{B.16a}$$

$$\hat{r}_N = \frac{r_N}{K} = \frac{N}{J_L} \left[(1 - m) \left(\frac{J_L - N}{J_L - 1} \right) + m(1 - p_k) \right]. \tag{B.16b}$$

Using (B.12), Eq. (5) becomes

$$\begin{aligned}
 P[N_k | J_L] &= P_k[0] \frac{J_L I p_k}{n_k [J_L - N_k + I(1 - p_k)]} \\
 &\quad \times \prod_{i=1}^{N_k-1} \frac{[J_L - i][i + I p_k]}{i [J_L - i + I(1 - p_k)]} \\
 &= P_k[0] \binom{J_L}{N_k} \frac{(I p_k)_{N_k}}{(I(1 - p_k) + J_L - N_k)_{N_k}} \\
 &= P_k[0] \binom{J_L}{N_k} \frac{(I p_k)_{N_k} (I(1 - p_k))_{J_L - N_k}}{(I(1 - p_k))_{J_L}} \\
 &= \binom{J_L}{N_k} \frac{(I p_k)_{N_k} (I(1 - p_k))_{J_L - N_k}}{(I)_{J_L}}. \tag{B.17}
 \end{aligned}$$

To calculate $P_k[0]$ we have used Vandermonde's formula, also known as the Chu–Vandermonde identity (Boros and Moll, 2004),

$$(x + y)_n = \sum_{n=0}^n \binom{N}{n} (x)_n (y)_{N-n}. \tag{B.18}$$

Eq. (B.17) was given by Vallade and Houchmandzadeh (2003).

Each term in (11) requires the probability of abundance n_k where we exclude the species $i = 1 \dots k - 1$. Thus, we rescale all relative abundances in the metacommunity,

$$\hat{p}_k = \frac{p_k}{1 - \sum_{i=1}^{k-1} p_i} \tag{B.19}$$

and the dispersal number becomes the effective dispersal number (see also Etienne and Alonso, 2005),

$$\hat{I}_k := I \left(1 - \sum_{i=1}^{k-1} p_i \right) = I \prod_{i=1}^{k-1} (1 - \hat{p}_i). \tag{B.20}$$

With this reparametrization, the situation for species k is identical to the situation for the first species:

$$\begin{aligned}
 & P[N_k | I_k, J_{L,k}, \hat{p}_k] \\
 &= \binom{J_{L,k}}{N_k} \frac{(\hat{I}_k \hat{p}_k)_{N_k} (\hat{I}_k (1 - \hat{p}_k))_{J_{L,k} - N_k}}{(\hat{I})_{J_{L,k}}}. \tag{B.21}
 \end{aligned}$$

Inserting this in (11), i.e. applying again the subsample approach, we obtain

$$\begin{aligned}
 P[n_1, \dots, n_S | I, J_L, \vec{p}] &= \prod_{k=1}^S P[N_k | \hat{I}_k, J_{L,k}, \hat{p}_k] \\
 &= \prod_{k=1}^S \binom{J_{L,k}}{N_k} \frac{(\hat{I}_k \hat{p}_k)_{N_k} (\hat{I}_k (1 - \hat{p}_k))_{J_{L,k} - N_k}}{(\hat{I}_k)_{J_{L,k}}} \\
 &= \prod_{k=1}^S \frac{J_{L,k}!}{N_k! (J_{L,k+1})!} \frac{(\hat{I}_k \hat{p}_k)_{N_k} (\hat{I}_k)_{J_{L,k+1}}}{(\hat{I}_k)_{J_{L,k}}} \\
 &= \frac{J_L!}{(I)_{J_L}} \prod_{k=1}^S \frac{(\hat{I}_k \hat{p}_k)_{N_k}}{N_k!} = \frac{J_L!}{(I)_{J_L}} \prod_{k=1}^S \frac{(I p_k)_{N_k}}{N_k!}. \quad (\text{B.22})
 \end{aligned}$$

Eq. (B.22) is invariant under sampling and thus applies to samples by replacing J_L by the sample size J . Because Eq. (B.22) is identical to (26), the zero-sum and non-zero-sum equilibrium species abundance distributions of the neutral model with dispersal limitation are identical.

References

- Alonso, D., McKane, A.J., 2004. Sampling Hubbell's neutral theory of biodiversity. *Ecol. Lett.* 7, 901–910.
- Alonso, D., Pascual, M., 2006. Comment on "A keystone mutualism drives pattern in a power function". *Science* 313, 1739b.
- Alonso, D., Etienne, R.S., McKane, A.J., 2006. The merits of neutral theory. *Trends Ecol. Evol.* 21, 451–457.
- Bascompte, J., Possingham, H., Roughgarden, J., 2002. Patchy populations in stochastic environments: critical number of patches for persistence. *Am. Natur.* 159, 128–137.
- Bascompte, J., Jordano, P., Melin, C.J., Olesen, J.M., 2003. The nested assembly of plant-animal mutualistic networks. *Proc. Nat. Acad. Sci. USA* 100, 9383–9387.
- Bell, G., 2005. The co-distribution of species in relation to the neutral theory of community ecology. *Ecology* 86, 1757–1770.
- Boros, G., Moll, V., 2004. Irresistible Integrals: Symbolics, Analysis and Experiments in the Evaluation of Integrals. Cambridge University Press, Cambridge, UK.
- Caswell, H., 1976. Community structure: a neutral model analysis. *Ecol. Monogr.* 46, 327–354.
- Chave, J., 2004. Neutral theory and community ecology. *Ecol. Lett.* 7, 241–253.
- Chave, J., Leigh, E.G., 2002. A spatially explicit neutral model of β -diversity in tropical forests. *Theor. Popul. Biol.* 62, 153–168.
- Chave, J., Alonso, D., Etienne, R.S., 2006. Comparing models of species abundance. *Nature* 441, E1–E2.
- Clinchy, M., Haydon, D.T., Smith, A.T., 2002. Pattern does not equal process: What does patch occupancy really tell us about metapopulation dynamics? *Am. Natur.* 159, 351–362.
- Cohen, J.E., 1968. Alternate derivations of a species-abundance relation. *Am. Natur.* 102, 165–172.
- Cohen, J.E., Jonsson, T., Carpenter, S.R., 2003. Ecological community description using the food web, species abundance, and body size. *Proc. Natl Acad. Sci. USA* 100, 1781–1786.
- Costantini, D., Garibaldi, U., 2004. The Ehrenfest fleas: from model to theory. *Synthese* 139, 107–142.
- Dewdney, A.K., 1998. A general theory of the sampling process with applications to the "veil line". *Theor. Popul. Biol.* 54, 294–302.
- Dewdney, A.K., 2000. A dynamical model of communities and a new species-abundance distribution. *Biol. Bull.* 35, 152–165.
- Diserud, O.H., Engen, S., 2000. A general and dynamic species abundance model, embracing the lognormal and the gamma models. *Am. Natur.* 155, 497–511.
- Dornelas, M., Connolly, S.R., Hughes, T.P., 2006. Coral reef diversity refutes the neutral theory of biodiversity. *Nature* 440, 80–82.
- Engen, S., Lande, R., 1996a. Population dynamic models generating the lognormal species abundance distribution. *Math. Biosci.* 132, 169–183.
- Engen, S., Lande, R., 1996b. Population dynamic models generating the species abundance distributions of the gamma type. *J. Theor. Biol.* 178, 325–331.
- Etienne, R.S., 2005. A new sampling formula for neutral biodiversity. *Ecol. Lett.* 8, 253–260.
- Etienne, R.S., 2007. A neutral sampling formula for multiple samples and an "exact" test of neutrality. *Ecol. Lett.* 10, 608–618.
- Etienne, R.S., Alonso, D., 2005. A dispersal-limited sampling theory for species and alleles. *Ecol. Lett.* 8, 1147–1156.
- Etienne, R.S., Alonso, D., 2006. Errata. *Ecol. Lett.* 9, 500.
- Etienne, R.S., Olf, H., 2004. A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.* 7, 170–175.
- Etienne, R.S., Olf, H., 2005. Confronting different models of community structure to species-abundance data: a Bayesian model comparison. *Ecol. Lett.* 8, 493–504.
- Etienne, R.S., Latimer, A.M., Silander, J.A., Cowling, R.M., 2006. Comment on "neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot". *Science* 311, 610b.
- Etienne, R.S., Apol, M.E.F., Olf, H., Weissing, F.J., 2007. Modes of speciation and the neutral theory of biodiversity. *Oikos* 116, 241–258.
- Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* 3, 87–112.
- Fisher, R.A., Corbet, A.S., Williams, C.B., 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* 12, 42–58.
- Gilbert, B., Laurance Jr., W.F., Nascimento, H.E.M., 2006. Can neutral theory predict the responses of Amazonian tree communities to forest fragmentation? *Am. Natur.* 168, 304–317.
- Goel, N.S., Richter-Dyn, N., 1974. *Stochastic Models in Biology*. Academic Press, New York, NY.
- Gotelli, N.J., Kelley, W.G., 1993. A general model of metapopulation dynamics. *Oikos* 68, 36–44.
- Harpole, W.S., Tilman, D., 2006. Non-neutral patterns of species abundance in grassland communities. *Ecol. Lett.* 9, 15–23.
- Harte, J., Kinzig, A., Green, J., 1999. Self-similarity in the distribution and abundance of species. *Science* 284, 334–336.
- Hairston, N.G., Ellner, S.P., Geber, M.A., Yoshida, T., Fox, J.A., 2005. Rapid evolution and the convergence of ecological and evolutionary time. *Ecol. Lett.* 8, 1114–1127.
- Holyoak, M., Leibold, M.A., Holt, R., 2005. *Metacommunities: Spatial Dynamics and Ecological Communities*. University of Chicago Press, Chicago, IL.
- Hoppe, F., 1987. The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol.* 25, 123–159.
- Hubbell, S.P., 1979. Tree dispersion, abundance and diversity in a tropical dry forest. *Science* 203, 1299–1309.
- Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.
- Karlin, S., McGregor, J., 1972. Addendum to a Paper of W. Ewens. *Theor. Popul. Biol.* 3, 113–116.
- Kendall, R.G., 1948. On some modes of population growth giving rise to R.A. Fisher's logarithmic series distribution. *Biometrika* 35, 6–15.
- Lande, R., Engen, S., Saether, B.-E., 2003. *Stochastic Population Dynamics in Ecology and Conservation*. Oxford University Press, Oxford, UK.
- Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M., Gonzalez, A., 2004. The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* 7, 601–613.
- MacArthur, R.H., 1957. On the relative abundance of bird species. *Proc. Natl Acad. Sci. USA* 43, 293–295.

- MacArthur, R.H., 1960. On the relative abundance of species. *Am. Natur.* 94, 25–36.
- MacArthur, R.H., Wilson, E.O., 1967. *Island Biogeography*. Princeton University Press, Princeton, NJ.
- Magurran, A.E., 2004. *Measuring Biological Diversity*. Blackwell, Oxford, UK.
- McGill, B.J., 2003. A test of the unified neutral theory of biodiversity. *Nature* 422, 881–885.
- McGill, B.J., Maurer, B.A., Weiser, M.D., 2006. Empirical evaluation of neutral theory. *Ecology* 87, 1411–1426.
- McKane, A.J., Alonso, D., Solé, R.V., 2000. Mean field stochastic theory for species rich assembled communities. *Phys. Rev. E* 62, 8466–8484.
- McKane, A.J., Alonso, D., Solé, R.V., 2004. Analytic solution of Hubbell's model of local community dynamics. *Theor. Popul. Biol.* 65, 67–73.
- Mosimann, J.E., 1962. On the compound multinomial distribution, the multivariate *beta*-distribution, and correlations among proportions. *Biometrika* 49, 65–82.
- Munoz, F., Couteron, P., Ramesh, B.R., Etienne, R.S., 2007. Inferring parameters of neutral communities: from one single large to several small samples. *Ecology*, in press.
- Nisbet, R.M., Gurney, W.S.C., 1982. *Modelling Fluctuating Populations*. Wiley, New York, NY.
- Pielou, E.C., 1969. *An Introduction to Mathematical Ecology*. Wiley, New York, NY.
- Pielou, E.C., 1975. *Ecological Diversity*. Wiley, New York, NY.
- Poulin, R., 2004. Parasites and the neutral theory of biodiversity. *Ecography* 27, 119–123.
- Preston, F.W., 1948. The commonness, and rarity, of species. *Ecology* 29, 254–283.
- Preston, F.W., 1962. The canonical distribution of commonness and rarity. Parts i and ii. *Ecology* 43, 185–215, 410–432.
- Pueyo, S., 2006. Diversity: between neutrality and structure. *Oikos* 112, 392–405.
- Purves, D.W., Pacala, S.W., 2005. Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. In: Burslem, D., Pinard, M., Hartley, S. (Eds.), *Biotic Interactions in the Tropics*. Cambridge University Press, Cambridge, UK, pp. 107–138.
- Rannala, B., 1996. The sampling theory of neutral alleles in an island population of fluctuating size. *Theor. Popul. Biol.* 50, 91–104.
- Renshaw, E., 1991. *Modelling Biological Populations in Space and Time*, Cambridge Studies in Mathematical Biology, vol. 11. Cambridge University Press, Cambridge, UK.
- Sugihara, G., 1980. Minimal community structure: an explanation of species-abundance patterns. *Am. Natur.* 116, 770–787.
- Tavaré, S., Ewens, W.J., 1997. Multivariate Ewens distribution. In: Johnson, N., Kotz, S., Balakrishnan, N. (Eds.), *Discrete Multivariate Distributions*. Wiley, New York, NY, pp. 232–246.
- Tokeshi, M., 1990. Niche apportionment or random assortment—species abundance patterns revisited. *J. Anim. Ecol.* 59, 1129–1146.
- Tokeshi, M., 1993. Species abundance patterns and community structure. *Adv. Ecol. Res.* 24, 111–186.
- Tokeshi, M., 1996. Power fraction: a new explanation of relative abundance patterns in species-rich assemblages. *Oikos* 75, 543–550.
- Turnbull, L.A., Manley, L., Rees, M., 2005. Niches, rather than neutrality, structure a grassland pioneer guild. *Proc. Roy. Soc. London B* 272, 1357–1364.
- Vallade, M., Houchmandzadeh, B., 2003. Analytical solution of a neutral model of biodiversity. *Phys. Rev. E* 68, 061902.
- Vallade, M., Houchmandzadeh, B., 2006. Species abundance distribution and population dynamics in a two-community model of neutral ecology. *Phys. Rev. E* 74, 051914.
- Van Kampen, N.G., 1992. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, The Netherlands.
- Volkov, I., Banavar, J.R., Hubbell, S.P., Maritan, A., 2003. Neutral theory and relative species abundance in ecology. *Nature* 424, 1035–1037.
- Volkov, I., Banavar, J.R., He, F., Hubbell, S.P., Maritan, A., 2005. Density dependence explains tree species abundance and diversity in tropical forests. *Nature* 438, 658–661.
- Wright, S., 1969. *Evolution and Genetics of Populations. The theory of gene frequencies*. University of Chicago Press, Chicago, IL.