Ecological Informatics 33 (2016) 101–108

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

A probability distribution model of small -scale species richness in plant communities

Jun Chen [a,*], Norbert Gáborčík [b], Masae Shiyomi [c]

[a] College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Dumbierská 32, 97411 Banská Bystrica, Slovakia
[c] Ibaraki Study Center, The Open University of Japan, Bunkyo 2-2, Mito 310-0056, Japan

ARTICLE INFO

Article history:
Received 12 July 2015
Received in revised form 8 April 2016
Accepted 10 April 2016
Available online 16 April 2016

Keywords:
Extended Poisson distribution
Grassland vegetation
Number of species per quadrat
Plant community
Species richness

ABSTRACT

We devised a probability distribution model that best expressed species richness per quadrat in grassland communities, and clarified the mechanism by which the mean richness per quadrat was always larger than the variance among quadrats. Our model will aid in the understanding of community structures, and allow comparisons among different communities. The model was constructed based on relatively simple theoretical assumptions about the mechanisms in play in target communities. We assumed in the model that the number of species occurring in an actual quadrat, j, is the sum of "the fundamental number of species", k (constant), and "a fluctuating number of species", i (a Poisson variate with the mean of μ); that is, $j = k + i$, where i, j and k are non-negative integers. The probability that j species occur in a quadrat is given by a Poisson-like distribution (extended Poisson), with two parameters k and μ. The mean species richness in the probability distribution is expressed by λ ($=k + \mu$), and the variance is $\lambda - k$. The proposed model afforded a good fit for the observed frequency distribution of species richness per quadrat. If even one species is common among many quadrats, the mean number of species per quadrat is greater than the variance. The greater the number of common species among quadrats is, the larger is the value of k, and then the more pronounced is the difference between the mean and the variance (although the variance does not change). We fitted the model to 55 datasets collected by ourselves from grasslands in various locations (Tibet, Inner Mongolia, Slovakia, or Japan), with varying quadrat size (0.25, 0.0625, or 0.01 m^2), and under differing management status (various stocking densities).

1. Introduction

Any herbaceous community that has a conceptually uniform structure is still spatially heterogeneous with respect to species richness, species composition, cover and biomass within the community when we look at a small scale (Chen et al., 2008b; Shiyomi et al., 2010). This heterogeneity can be caused by the local disappearance of species from the community, local colonization of new species from neighboring communities, intra- and inter-specific interactions, and chemical, physical, and biological disturbances. Additionally, anthropogenic disturbances, such as feeding and trampling by livestock, weeding, and fertilization, accelerate spatial heterogeneity in these communities (Shiyomi et al., 2010; Whittaker and Naveh, 1979).

In this study, we enumerated the species in multiple small quadrats in a community through field observations. Spatial heterogeneity in such communities is often found in grasslands (e.g., Kull and Zobel, 1991; Duncan et al., 1998; Zobel et al., 2000; Chen et al., 2008a, 2008b). For example, the literature contains much discussion regarding small-scale spatial heterogeneity in species richness, such as whether it is caused by small-scale niche division, disturbances or accidents, or temporary colonization from surrounding environments (e.g. Hubbell, 2001; Lavorel et al., 1994; Schmida and Wilson, 1985; van der Maarel and Sykes, 1993; van der Maarel et al., 1995; Wilson et al., 1995).

By setting multiple small quadrats (e.g., 50 × 50 cm) in a grassland, counting the number of species, and summarizing them in a frequency distribution, we can usually obtain a symmetric discrete distribution, with the mean at approximately the center of the distribution (e.g. Chen et al., 2005; Shiyomi et al., 2004; Tsutsumi et al., 2003). However, whether we can express this frequency distribution in a mathematical form with a simple, biologically significant meaning has been a longstanding question. If each species within a community occurs randomly among quadrats, the expected frequency distribution (null frequency distribution) could be expressed by a simple probability distribution (Shiyomi et al., 2010). The null frequency distribution is used for testing whether the actual frequency distribution of species richness obtained via a field survey follows a random distribution. However, the null frequency distribution may not be the actual frequency distribution found in grasslands. The primary objective of this study was to find a theoretical probability distribution model that can be fitted to the observed frequency distributions of species richness for any

* Corresponding author.
E-mail address: chen_jun2008@nwsuaf.edu.cn (J. Chen).

http://dx.doi.org/10.1016/j.ecoinf.2016.04.003
1574-9541/

actual plant community, rather than the null frequency distribution or empirical frequency/probability distributions. The frequency distribution of species richness as well as species composition can be used to better understand the structure of plant communities and to make comparisons between different plant communities, e.g., under different grassland management practices and in different landscapes.

Various studies have determined the mean species richness of plant communities in different landscapes. For example, Zobel et al. (2000) surveyed two species-rich alvar grasslands characterized by *Filipendula hexapetala–Trifolium montanum* in Hanila, Estonia, using sixty 10 × 10-cm quadrats. At the first site, the mean was 13.2 species per quadrat and the variance was 2.9 (calculated based on the standard error shown in their paper), while at the second site, the mean was 17.1 species per quadrat and the variance was 3.17. Likewise, Kull and Zobel (1991) surveyed two communities in a wooded meadow characterized by *Sesleria coerulea–F. hexapetala* in Laelatu, Estonia, using 30 quadrats 10 × 10 cm in size. The first community contained 4.0 species per quadrat with a variance of 1.45, and the second contained 17.7 species per quadrat with a variance of 7.20. In 2005, Chen et al. (2008b) surveyed two semi-natural grasslands characterized by *Zoysia japonica* with different stocking rates in central Japan using 100 quadrats of 10 × 10 cm. They found that in a heavily grazed pasture, the mean was 4.75 species and the variance was 2.27, whereas in a lightly grazed pasture, the species richness was 4.4 and the variance was 2.65. Kent (2012) produced a dataset for Gutter Tor, Dartmoor, in southwest England, where he surveyed 25 quadrats measuring 5 × 5 m and found a total of 27 species. The mean richness per quadrat and variance were 5.92 and 2.08, respectively. Finally, in 2003, Chen et al. (2008c) surveyed a grazed alpine meadow (3200 m a.s.l.) characterized by *Kobresia humilis* in Haibei (Qinghai, China), using 100 quadrats 10 × 10 cm in size, and found 19.69 species per quadrat with a variance of 14.64. In all of the above examples, the mean species richness per quadrat was larger than the variance. Therefore, a further aim of this study was to determine, through the model, why the mean number of species per quadrat is generally larger than the variance in the number of species among quadrats. The probability distribution model, which is described in the following sections, was characterized using the mean and variance of species richness among quadrats.

2. Model

Suppose that we set N small quadrats (for example, 100), each of which has a given area, in a uniform plant community in a uniform landscape. For any species, the frequency of occurrence per quadrat is dependent on the spatial abundance of the species throughout the community. In a grassland, if each species in a community occurs in random quadrats, species with a very high occurrence often occurs in most or all quadrats set in the community, but species with a low occurrence does only in a few quadrats. We assumed that many species, with varying levels of occurrence, coexist in each quadrat, and all of the species occur in arbitrary quadrats. Furthermore, we assumed that the number of species in each quadrat is equal or larger than a constant, non-negative integer k. It may be more convenient to consider that k is a non-negative rational number, as shown in the Appendix. A probable ecological concept is that, in a uniform community, a given number of species (k) is first arranged at each quadrat, and an additional number of species, which is a random number, is added to the quadrat (an example is shown below). The value of k is strongly affected by species with high occurrence, but it is also somewhat affected by species with a low occurrence. The value k indicates the fundamental number of species per quadrat. The value of k is also dependent on the quadrat size, with a large value of k found in a large quadrat.

We assumed on the above concept that the number of species that occurred in an actual quadrat, j, is the sum of "the fundamental number of species (referred to as Fus)", k (constant), and "a fluctuating number of species (Fls)", i (variate):

$$j = k + i, \tag{1}$$

where $i = 0, 1, 2, \ldots$ As the probability distribution for i, we assumed a Poisson distribution with mean μ (therefore, the variance is also μ) because i seems to be a rare event. Thus, the probability that j species occur in a quadrat (or the relative frequency of quadrats in which j species occur when many quadrats are surveyed), $P(j)$, is determined by the following equations:

$$P(0) = P(1) = \ldots = P(k-1) = 0, \tag{2a}$$

$$P(j) = e^{-\mu}\mu^{j-k}/(j-k)!, \text{for } j = k+i, i = 0, 1, 2, \ldots. \tag{2b}$$

Let λ and σ^2 be the mean and variance of $P(j)$, respectively. If $k = 0$, Eq. (2a) vanishes, and Eq. (2b) is the usual Poisson distribution, where i is the Poisson variate, and μ is both the mean and the variance. We expect the following additive relationship based on Eq. (1):

$$\lambda = k + \mu. \tag{3}$$

In some cases, Eq. (3) may be replaced by the relation of "λ nearly equal to $k + \mu$" as shown in the Appendix.

Eqs. (2a) and (2b) are the shifted function of the usual Poisson distribution to the positive direction by k units (an example will be shown later). In this model, because the variance of variate j is equal to the variance of i, Eq. (3) can also be written as: $\lambda = \sigma^2 + k$ ($k \geq 0$). Therefore, we obtain the relation of $\lambda \geq \sigma^2$. We hereafter refer to (2a) and (2b) as the "extended Poisson distribution" or briefly "E Poisson".

We offer an actual example. We set 90 quadrats, each of which had an area of 50 × 50 cm, in a community of semiarid natural grassland (Shenmu, Shaanxi, China) and counted the number of species in each quadrat (data originated from Lv et al., 2011). The data obtained are shown in Fig. 1a. In this case, the numbers of species per quadrat was ≥4 in all quadrats, i.e., $k = 4$ (^ indicates the estimated value; Fus is 4). The number of quadrats with 0–3 species is 0, the number of quadrats that contained four species was 5, the number of quadrats that contained six species was 14, and so on. The quadrat-to-quadrat data were summarized in a frequency distribution as shown in Fig. 1b. The estimated mean number of species per quadrat, λ, and the estimated variance of species among quadrats, σ^2, were 6.76 and 2.23, respectively (therefore, the mean > the variance). The estimated mean, μ, was 2.76 from Eq. (3), and the estimated variance of i was 2.23. The difference between the mean, μ, and variance, σ^2, is 0.53. This small difference indicates a high possibility that the frequency distribution of i can be properly approximated by a Poisson distribution with a mean of 2.76. This is an experimental evidence of the model concept.

The spatial heterogeneity is determined by the variance/mean ratio, i.e., $I = \sigma^2 / \lambda$ (Devid and Moore, 1954). For the ordinary Poisson distribution, i.e., random distribution, $I = 1$; for heterogeneity lower than would be expected by a random distribution, $I < 1$; and for heterogeneity greater than would be expected by a random distribution, $I > 1$. The larger the I value is, the greater the spatial heterogeneity of the number of species becomes. In the above case, $I = 0.33$, which indicates a highly uniform pattern relative to the random distribution.

The calculated frequency distribution of the number of species per quadrat, $E(j)$, based on $k = 4$ and $\mu = 2.76$ is shown in the right-hand side of Fig. 1c, where $E(j) = P(j) \times N$, and $N = 90$ (Eqs. (2a) and (2b)). Whether the model is right is determined by the goodness-of-fit test based on the chi-square test. In this case, the model is shown to fit the observed data because the probability that the null hypothesis (i.e., H_0: the model is equal to the observed frequency distribution) is accepted is 0.28, which is >0.05 (the calculated χ_0^2 was 5.06 for 4 degrees