

Analysis of diversity

Analysis of diversity

This chapter describes some methods of investigating diversity. First, the concept of diversity is introduced. Then some methods of calculating and comparing diversity are discussed.

Diversity entails richness (or the number of species) and evenness (or equality in the number of individuals for every species).

Where is diversity the highest?

In this chapter, some methods are introduced to calculate the diversity of a specific site (a sample

plot in a forest, a farm, a village, ...). Consider Figure 5.1, for instance. You may be interested in finding out whether the diversity is higher in site B than in site A. You could for instance have a hypothesis that the diversity in site B is greater because temperatures are higher in site B. This chapter will only describe the **methods of calculating the diversity** of a site. To test a hypothesis for the relationship between some explanatory factors of diversity and diversity of a site, you will need to use a regression method as described in chapter 6 and two sites will definitely not be a large enough sample size to investigate such hypothesis.

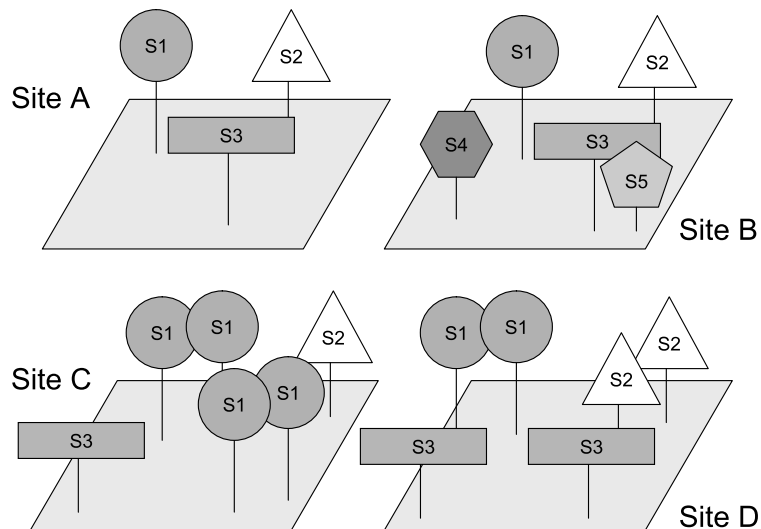


Figure 5.1 (a) Various sites differ in the number of species and in the number of trees of each species.

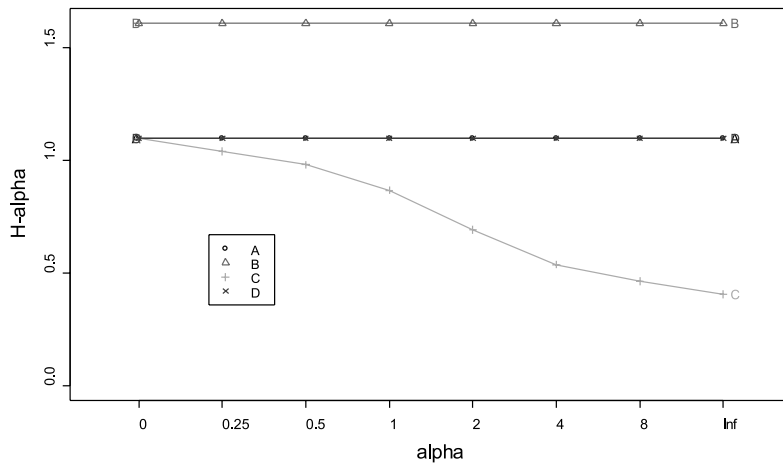


Figure 5.1 (b) Sites that are more diverse have a profile (the line in the diagram) that is higher, therefore the diversity ordering is: $B > A = D > C$.

What is diversity?

In general, diversity refers to the number of categories that can be differentiated, and to the proportions (or relative abundances) of the number of objects in each category. When we study tree species diversity, the categories refer to different species, whereas the objects are the trees that are counted.

Imagine that you have 2 sites: site A has 3 tree species, whereas site B has 5 tree species. In this situation, site B has the largest **species richness**. This situation is depicted in Figure 5.1.

Imagine another situation where both site C and site D contain 3 species. However, site C is dominated by one species that has 4 trees out of the total number of 6 trees on the entire site (or a proportion of $4/6$). The other two species have proportions of $1/6$. In site D, each species has the same number of trees (or proportions of $2/6$). In this situation, site D has the largest **evenness**, which means that the proportions of the individual species are more similar. In this situation, the proportions are actually all the same for site D, so evenness is maximum for this site. This situation is also shown in Figure 5.1.

Sites A and D (Figure 5.1) have the same proportions. Since both sites have the same proportions and the same number of species, they have the same diversity. Diversity does not depend on density or total abundance.

Sites of maximum evenness will have proportions of $1/S$ for each species, where S is the number of species (the species richness). For example, a plot with 5 species of maximum evenness will have proportions of $1/5$ for each species, whereas a farm with 10 species of maximum evenness will have proportions of $1/10$ for each species. If 100 trees were recorded in total, then 5 species will be most evenly distributed when each species has $100/5 = 20$ trees.

On the other hand, a site of minimum evenness will have only 1 tree for the $S-1$ less frequent species and $Tot - (S-1)$ trees for the dominant category, if Tot indicates the total number of trees. If a site contains 6 trees and 3 species, the minimum evenness will be where 2 species contain 1 tree and the remaining species contains $6-2=4$ trees. If a site contains 100 trees and 5 species, then with minimum evenness the dominant species will contain $100-4=96$ trees.

In most situations, the evenness will be in between the maximum and minimum evenness.

Since diversity refers to richness and evenness, both these facets need to be considered when comparing diversity. If evenness is the same for the sites (sites, farms, sample plots) that you are comparing, then differences in richness will correspond to differences in diversity. If the richness is the same, then differences in evenness will correspond to differences in diversity. There will be situations, however, where one site has larger richness but lower evenness than another site. In these situations, it is not always possible to rank one site as higher in diversity than the other site.

Rank-abundance curves

Rank-abundance curves are conceptually the easiest method of analysing patterns of diversity.

First, the total number of individuals is calculated for each species. Second, species are ranked from the most abundant to the least abundant. Finally a plot is constructed with the rank number on the horizontal axis, and the abundance on the vertical axis.

For example, the rank-abundance pattern for the dune meadow dataset (we treated the values in the cells of the species matrix as if they were counts of individuals) is:

	rank	abundance	proportion
Poatri	1	63	9.2
Lolper	2	58	8.5
Leoaut	3	54	7.9
Brarut	4	49	7.2
Agrsto	5	48	7.0
Poapra	6	48	7.0
Trirep	7	47	6.9
Alogen	8	36	5.3
Elyrep	9	26	3.8
Plalan	10	26	3.8
Elepal	11	25	3.6
Antodo	12	21	3.1
Sagpro	13	20	2.9
Junart	14	18	2.6
Rumace	15	18	2.6
Achmil	16	16	2.3
Brohor	17	15	2.2
Ranfla	18	14	2.0
Belper	19	13	1.9
Junbuf	20	13	1.9
Salrep	21	11	1.6
Calcus	22	10	1.5
Hyprad	23	9	1.3
Triptra	24	9	1.3
Airpra	25	5	0.7
Potpal	26	4	0.6
Viclat	27	4	0.6
Cirarv	28	2	0.3
Empnig	29	2	0.3
Chealb	30	1	0.1

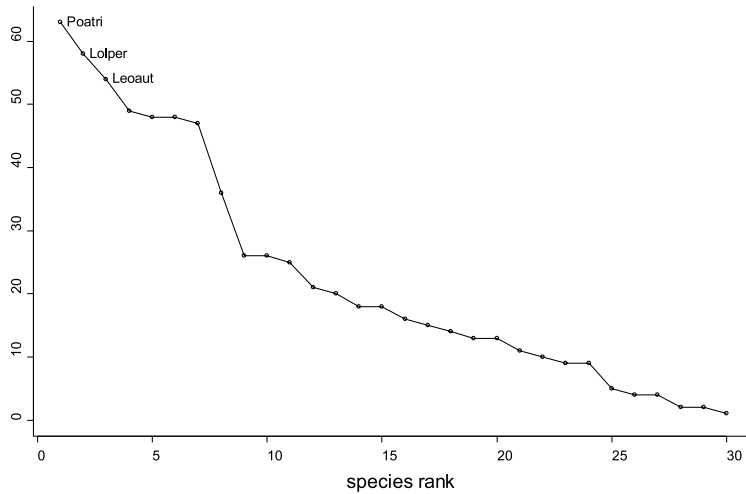


Figure 5.2 Rank-abundance curve for the dune meadow dataset.

The results given before are normally provided as a rank-abundance curve such as Figure 5.2.

You can see that *Poa trivialis* was ranked 1 as this species had the largest total abundance of 63, and that *Chenopodium album* was ranked 30 since this species had the lowest total abundance of 1. Figure 5.2 shows the rank-abundance curve for this dataset. You could create an alternative rank-abundance curve by plotting the proportion instead of the abundance on the vertical axis. The shape of the curve would remain the same, since only the scaling of the vertical axis would be different (note also that the proportion was given as percentage by multiplying all proportions with 100%). Other alternatives include plotting the logarithm of abundance on the vertical axis – this could produce better graphs when a few species are highly dominant.

The interpretation of a rank-abundance curve in terms of diversity, i.e. richness and evenness, is as follows. On the horizontal axis, species richness is provided by the width of the curve. A wider curve will indicate higher species richness. The shape of the rank-abundance curve

is an indication of the evenness. A completely horizontal curve is an indication of a completely evenly distributed system. The steeper the curve, the less evenly species are distributed.

Figure 5.3 provides the rank-abundance curves for the four sites shown in Figure 5.1. The proportion is plotted on the vertical axis. Note that sites A and D have the same rank-abundance curve when scaled to proportion (each species has proportion = 1/3). You can see from the widths of the rank-abundance curves of Figure 5.3 that one site has species richness of 5, whereas the other sites have richness 3. You can also see that three sites have completely horizontal profiles or completely evenly distributed species. You can notice one site with a declining profile, indicating that some species have higher abundance than others. In other words, species are not evenly distributed for this last site. Based on this information, you could classify site B as the most diverse (highest richness [=widest] and evenness [=most horizontal]), and site C as the least diverse (lowest richness [=narrowest] and evenness [=least horizontal]).

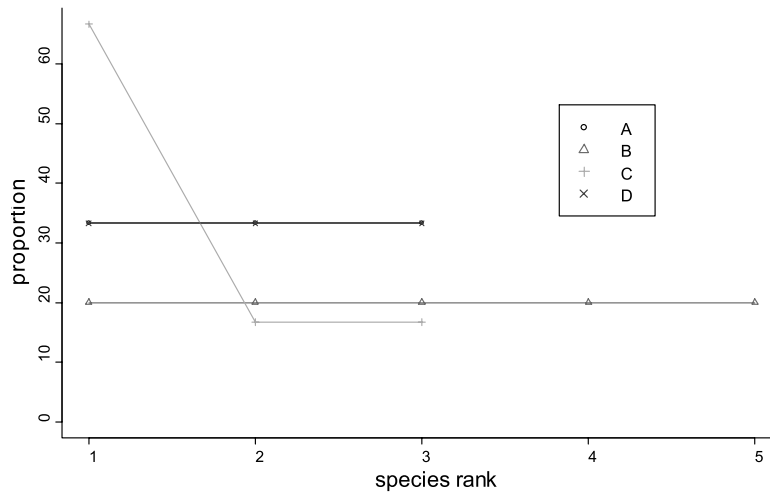


Figure 5.3 Rank-abundance curves for the 4 sites of Figure 5.1. Abundance is proportional abundance (percentage of each species of total abundance).

Models for rank-abundance curves

Various studies have been conducted to model specific rank-abundance distributions. By fitting a model, the shape of a particular rank-abundance distribution may be summarized by a few parameters. As some models are derived from theoretical assumptions about the ways in which species could coexist, the observation of a rank-abundance distribution that conforms to a particular model provides some evidence that the conditions that generate the model could apply to a particular survey. This could be important information since the question why species differ in abundance has been an important topic of biodiversity research.

A thorough discussion of rank-abundance distribution models is beyond the scope of this

manual. The interested reader could consult a specialized text such as Hubbell (2001). Models should not be fitted because it is possible, but because the information that they provide is useful. If you only want to summarize the rank-abundance distribution, you should reconsider whether you would not provide better information by just providing the actual rank-abundance curve.

Different models have been formulated to describe rank-abundance distributions, including the log-normal, log series and geometric distributions. Fitting these distributions to data is not difficult but it is often difficult to choose the one model that provides the best fit to the data.

When you attempt to fit several models to the dune meadow dataset, you will obtain following results:

```

RAD models, family poisson
No. of species 30, total abundance 685

Warning: NAs introduced by coercion
      par1   par2   par3 Deviance    AIC    BIC
Preemption 0.09674                16.852 155.570 156.972
Lognormal   3         1                38.217 178.936 181.738
Veiled.LN   3         1         1        38.217 180.936 185.140
Zipf        0.14817       -1        106.934 247.653 250.455
Mandelbrot  Inf -463703 4621740    15.204 157.923 162.127

```

It is easy to compare the fits graphically in this example as shown in Figure 5.4. Figure 5.4 shows the fit of the various models. Visually comparing the difference in the actual values and the predicted values allows you to choose a model that might fit your purpose. Of those in Figure 5.4, none capture the curvature of the observed distribution at high and particularly the low abundance end of the distribution. Over much of the range the veiled lognormal curve seems to fit best. Note that a logarithmic scale was used on the vertical axis.

An additional method of choosing the best distribution that you could use together with the graphical evaluation is to choose the model with the lowest AIC or BIC (Akaike's Information Criterion or the Bayesian Information Criterion; these are statistics that indicate goodness-of-fit of a model – lower values indicate better fits). For the dune meadow dataset, we might thus prefer the pre-emption model. But note that the statistics such as deviance, AIC and BIC only measure some aspects of fit, and they might not be the aspects you are most interested in.

Rényi diversity profiles

Rényi diversity profiles are curves that also provide information on richness and evenness, as rank-abundance curves do. Rényi diversity profiles have the advantage over rank-abundance curves that ordering from lowest to highest diversity is easier. For this reason, a Rényi diversity profile is one of several **diversity ordering techniques** (Tóthmérész 1995). The disadvantage of these curves is that information on the proportions of each species is not provided any longer.

Figure 5.1b provides the Rényi profiles for the same sites shown in Figure 5.3. The interpretation of a profile is as follows. The shape of the profile is an indication of the evenness. A horizontal profile indicates that all species have the same evenness – the same situation as for rank-abundance curves. The less horizontal a profile is, the less evenly species are distributed. In Figure 5.1b, we see that 3 sites have horizontal profiles, which means that species are completely evenly distributed for these

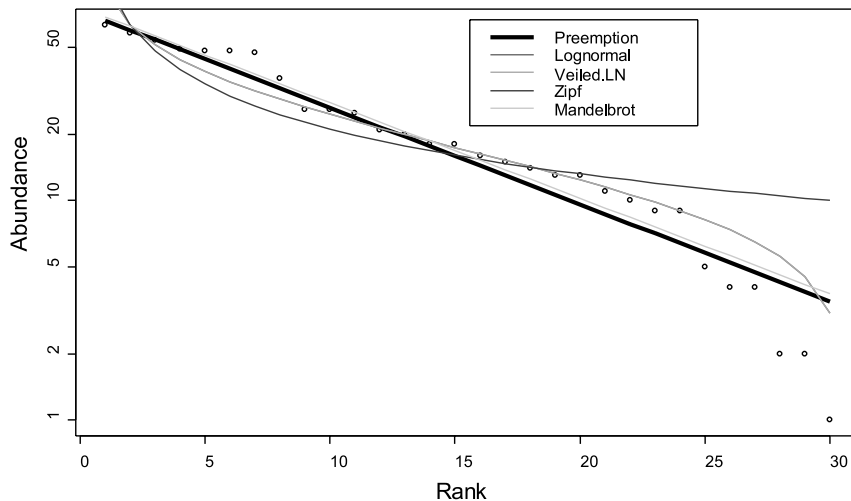


Figure 5.4 Fits of various models to the rank-abundance distribution of the dune meadow dataset.

sites. Site C has a profile that declines from left to right. This indicates that species are not evenly distributed for this site. The starting position at the left-hand side of the profile is an indication of the species richness. Profiles that start at a higher level have higher richness.

The major advantage of Rényi diversity profiles is that sites can easily be ordered from high to low diversity. If the profile for one site is everywhere above the profile for another site, then this means that the site with the highest profile is the more diverse of the two. From Figure 5.1b, you can for instance see that site B is the most diverse, and site C is the least diverse. If the profiles intersect, it is not possible to order the sites from lowest to highest diversity. It is possible that one site has larger species richness, but lower species evenness, although this is not a necessary condition for intersecting profiles.

If one diversity profile is higher than another, then the corresponding cumulated proportions of the rank-abundance curve will be lower. The cumulated proportions are calculated as the sum of the proportions of the rank-abundance curve

for 1, 2, 3, ..., all species. Figure 5.5 shows the cumulated proportions for the rank-abundance curves of Figure 5.3. The curve that is lowest everywhere in the figure corresponds to the site with the highest diversity – in this example, this is site B. You can verify that the ranking of diversity that is portrayed in Figure 5.5 ($B > A = D > C$) is the same as the ranking of Figure 5.1b.

Since the shape of the Rényi diversity profiles is influenced by evenness, you can compare the evenness of various sites by only looking at the shape of these curves. **Rényi evenness profiles** are a more direct method of comparing evenness (Ricotta 2003, Kindt et al. in press). These evenness profiles only reflect differences in evenness. The way that the evenness profiles are interpreted is similar to the way that diversity profiles should be interpreted, and the only difference is that a graphical comparison of evenness rather than of diversity is provided. An area of larger evenness will have an evenness profile that is everywhere above the evenness profile of an area of lower evenness. Intersecting evenness profiles means that no ranking in evenness can be provided. Figure 5.6

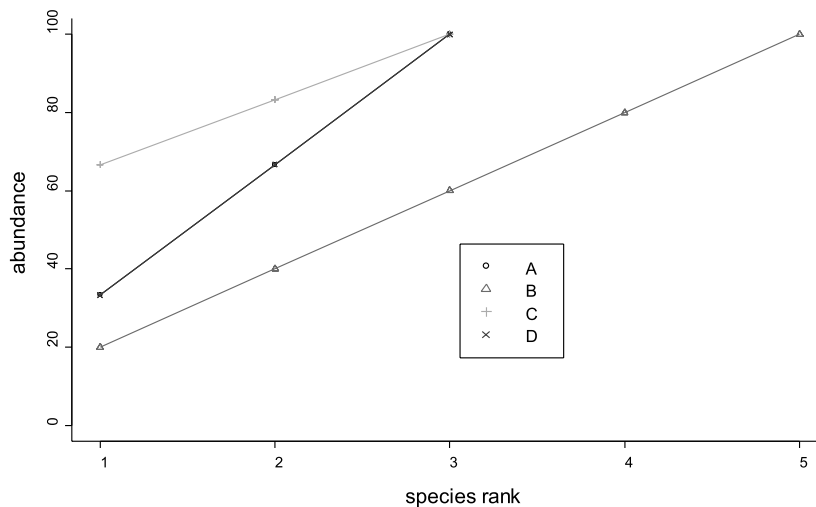


Figure 5.5 Cumulated proportions (%) for the 4 sites of Figure 5.1. This is an alternative diversity ordering technique to the Rényi diversity profile, with the lowest curve over the entire range indicating the site of highest diversity.

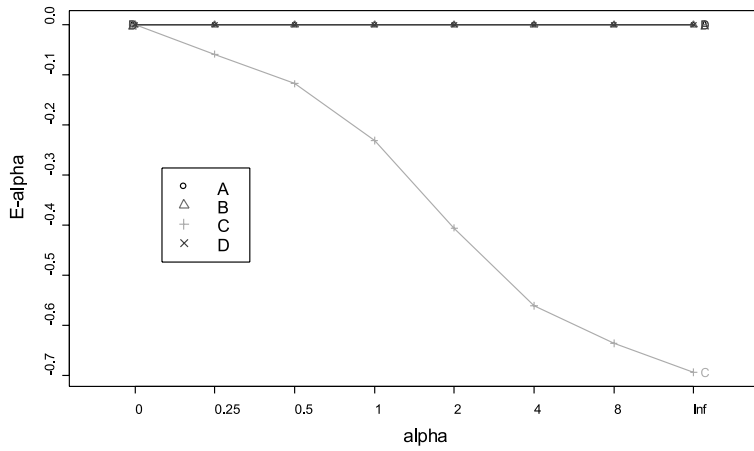


Figure 5.6 Rényi evenness profiles for the 4 sites of Figure 5.1.

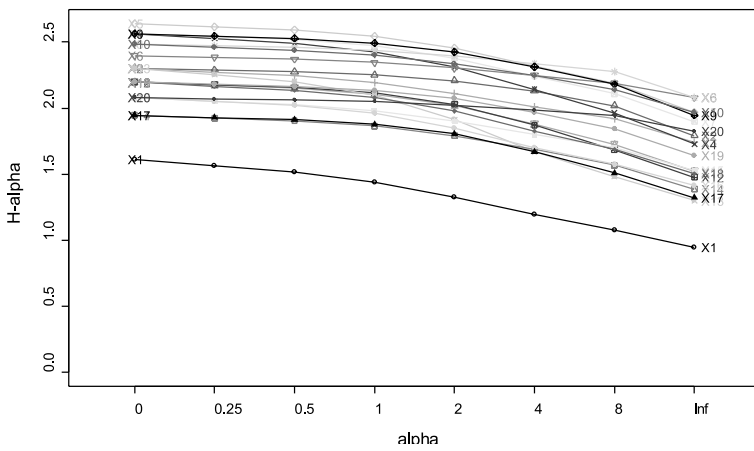


Figure 5.7 Rényi diversity profiles for the separate sites of the dune meadow dataset.

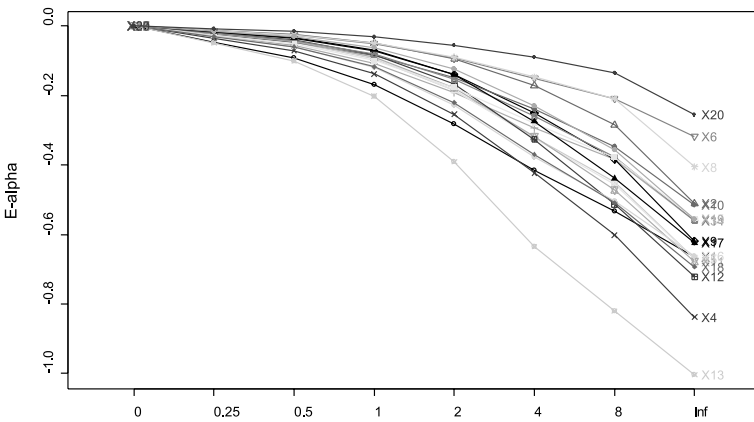


Figure 5.8 Rényi evenness profiles for the separate sites of the dune meadow dataset.

Each value of the Rényi diversity profile is based on a parameter 'alpha'. Box 5.1 shows how the profile value is calculated from the species proportions and the alpha parameter.

The profile values for alpha=0 provide information on species richness. The profile value is the logarithm of the species richness. For site A, $1.098612 = \ln(3)$. Thus, if you take the anti-logarithm ($y=\exp(x)$) of a Rényi diversity profile, you will obtain the species richness at alpha=0. For site B, species richness = $5 = \exp(1.609438)$. This is the reason that profiles that start at a higher level correspond to sites that are richer.

The profile value for alpha=infinity provides information on the proportion of the most abundant species. The profile value for

alpha=infinity for system C equals 0.4054651. When you take the anti-logarithm ($y=\exp(x)$), and then take the reciprocal value, then you will obtain the proportion of the most dominant species. For system C, the proportion of the most dominant species = $4/6 = 1/\exp(0.4054651)$. As a consequence, profiles that are higher at alpha=infinity have a lower proportion of the dominant species. A larger evenness thus corresponds with lower proportions of the dominant species.

The profile value for alpha=1 is the Shannon diversity index (discussed in the next section). The profile value for alpha=2 is the logarithm of the reciprocal Simpson diversity index (see next section).

Box 5.1 How to calculate the Rényi diversity profile?

The formula to calculate the diversity profile is:

$$H_{\alpha} = \frac{\ln\left(\sum_{i=1}^S p_i^{\alpha}\right)}{1-\alpha}$$

The p_i values of the above formula are the proportions of each species. For site A (Figure 1), these are $p_1 = p_2 = p_3 = 1/3$.

To calculate the profile value for $\alpha=8$, these proportions are raised to power α ($= 8$) = $(1/3)^8 = 0.000152$ and added together for all species as indicated by the summation symbol α ($0.000152 + 0.000152 + 0.000152 = 0.000457$). Next the logarithm is taken ($\ln(0.000457) = -7.690$). Finally, the obtained value is divided by $(1-\alpha = -7)$ ($-7.690/-7 = 1.0986$).

A profile is calculated by changing the value of α from 0 to infinity. In BiodiversityR, the standard values for α are: 0, 0.25, 0.5, 1, 2, 4, 8, and infinity. Although the formula can not be used directly to calculate the profile value at scales 1 and infinity, other formulas can be used to calculate these values since it is known that these values are related to the Shannon and Berger-Parker diversity indices (see main text). You can directly calculate the diversity profile by the `renyi` function (see at the end of this chapter).

Diversity indices

Diversity indices provide a summary of richness and evenness by combining these two facets of diversity into a single statistic. There are many ways by which richness and evenness can be combined, and this has resulted in many different diversity indices. Some of the common diversity indices are the Shannon, Simpson, and log series alpha diversity indices. A larger Simpson index will indicate lower diversity, hence it is better to analyse the reciprocal value of the Simpson index. An alternative approach is to report 1-Simpson index.

For the dune meadow dataset, you will obtain following results for the Shannon and 1-Simpson diversity indices (for each diversity index, the sites are sorted in increasing order) (shown on the right-hand side):

Diversity indices are a more compact method of comparing diversity. However, one diversity index will often not provide sufficient information to order sites from high to low diversity. Only diversity ordering techniques such as the Rényi diversity profiles will provide enough information that will allow you to conclude that one site is more diverse than another site. The reason for this phenomenon is actually that not all entities can be ordered from lowest to highest diversity (as shown earlier for the dune meadow dataset, see Figure 5.7).

It is true that if site A is more diverse than site B, that then the diversity index of site A will be larger. It is however not necessarily true that if the diversity index of site A is larger than the diversity index of site B, that then the diversity of site A is larger.

You could see in this dataset that the Shannon index of site X13 > X20, but that for the Simpson index X13 < X20. This is one illustration of the fact that a single diversity index may not provide sufficient information for diversity ordering. In Figure 5.7, you can see the intersection in the profiles of X13 and X20.

Shannon	
X1	1.440
X14	1.864
X17	1.876
X16	1.960
X15	1.979
X20	2.048
X18	2.079
X13	2.100
X11	2.106
X12	2.114
X19	2.134
X3	2.194
X2	2.253
X6	2.346
X10	2.399
X4	2.427
X8	2.435
X7	2.472
X9	2.494
X5	2.544
inverseSimpson	
X1	3.767
X14	6.000
X17	6.081
X16	6.368
X15	6.696
X13	6.764
X18	7.218
X11	7.529
X20	7.567
X12	7.609
X19	7.942
X3	8.247
X2	9.093
X6	10.017
X4	10.075
X10	10.330
X7	10.811
X8	10.959
X9	11.308
X5	11.629

Comparing the total diversity of different subsets of the dataset

Similarly to comparisons of species richness, you need to be cautious if you want to compare the total diversity of various subsets in your data when these subsets have different sample sizes. As for species richness, diversity indices will also change when sample size is increased. This is to be expected since diversity indices provide information on richness and evenness – so if richness changes with sample size, then diversity will change too.

If you want to compare the total diversity of subsets in your data, then you need to calculate the diversity for subsets in your data that have the same sample size. A procedure similar to the randomisation approach discussed for species accumulation curves

can be used. This procedure involves taking random subsets of the data and calculating a diversity profile for the subset. By randomized resampling of the subsets, average values of the diversity profiles can be obtained. Figure 5.9 shows an accumulation pattern for the average Rényi diversity profile for the dune meadow dataset.

If we want to compare the diversity of the different management categories of the dune meadow dataset for example, then we need to compare the diversity at the same sample size for the various categories. In this dataset, the largest sample size at which this is possible is 3, the number of sites of hobby farming. The results of a comparison at sample size 3 for the dune meadow dataset is shown in Figure 5.10. From this figure, you could conclude that hobby farming is the

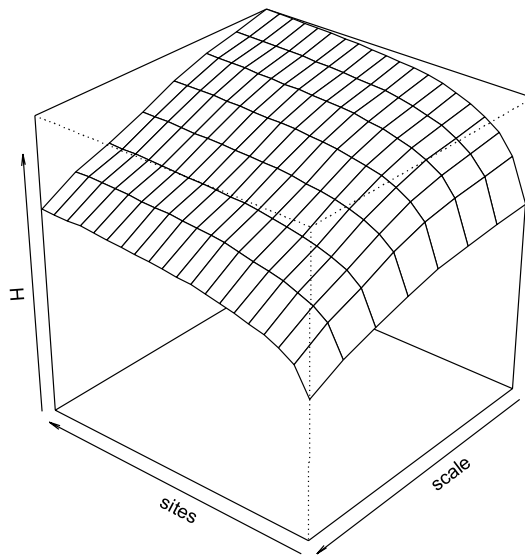


Figure 5.9 Accumulation pattern for the average Rényi diversity profiles for the dune meadow dataset.

most diverse category when comparing average diversity profiles for combinations of 3 sites.

As for species accumulation curves, there are various options to measure the same sample size. You could either choose the number of sites as a measure of sample size, or the number of plants, or the area that was sampled – and some other measures could theoretically be chosen too.

As we saw in the previous chapter, it is not necessarily so that hobby farming will be the most diverse at the scale of the entire landscape, although it is the most diverse at sample size 3. It could be possible that only 5% of the entire landscape is

under hobby farming whereas standard farming could be 80% of the landscape. In such situations, standard farming could be more diverse at the scale of the entire landscape. Since we have only sampled a fraction of the landscape and since extrapolation is difficult, we have no evidence that standard farming or hobby farming is more diverse at the landscape level. When your primary interest is to understand the distribution of biodiversity, then it may also be worthwhile to investigate differences in species composition (Chapter 8 and beyond) rather than differences in total diversity of different subsets in your data.

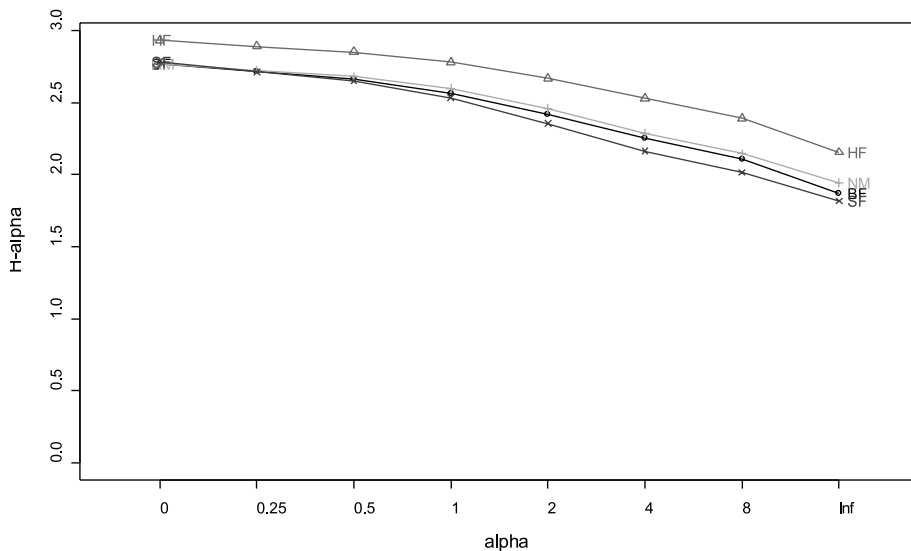


Figure 5.10 Comparison of diversity for the management categories of the dune meadow dataset. Results are based on 100 randomisations.

References

- Feinsinger P. 2001. *Designing field studies for biodiversity conservation*. Washington: The Nature Conservancy.
- Hayek L-AC and Buzas MA. 1997. *Surveying natural populations*. New York: Columbia University Press.
- He F and Legendre P. 2002. Species diversity patterns derived from species-area models. *Ecology* 83: 1185-1198.
- Hubble SP. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton: Princeton University Press.
- Kempton RA. 2002. Species diversity. In: El-Shaarawi, AH and Piegorosch WW. *Encyclopedia of environmetrics*. Chichester: John Wiley and Sons.
- Kent M and Coker P. 1992. *Vegetation description and analysis: a practical approach*. London: Belhaven Press.
- Kindt R, Van Damme P and Simons AJ. (in press). Tree diversity in western Kenya: using profiles to characterise richness and evenness. *Biodiversity and Conservation*.
- Krebs CJ. 1994. *Ecological methodology*. Second edition. Menlo Park: Benjamin Cummings.
- Legendre P and Legendre L. 1998. *Numerical ecology*. Amsterdam: Elsevier Science BV.
- Magnussen S and Boyle TJB. 1995. Estimating sample size for inference about the Shannon-Weaver and the Simpson indices of species diversity. *Forest Ecology and Management* 78: 71-84.
- Magurran AE. 1988. *Ecological diversity and its measurement*. Princeton, N.J: Princeton University Press. (recommended as first priority for reading)
- Magurran AE and Henderson PA. 2003. Explaining the excess of rare species in natural species abundance distributions. *Nature* 422: 714-716.
- Pielou EC. 1969. *An introduction to mathematical ecology*. New York: Wiley - Interscience.
- Purvis A and Hector A. 2000. Getting the measure of biodiversity. *Nature* 405: 212-218.
- Ricotta C. 2003. On parametric evenness measures. *Journal of Theoretical Biology* 222: 189-197.
- Rousseau D, Van Hecke P, Nijssen D, and Bogaert J. 1999. The relationship between diversity profiles, evenness and species richness based on partial ordering. *Environmental and Ecological Statistics* 6: 211-223.
- Shaw PJA. 2003. *Multivariate statistics for the environmental sciences*. London: Hodder Arnold.
- Tóthmérész B. 1995. Comparison of different methods for diversity ordering. *Journal of Vegetation Science* 6: 283-290.

Doing the analyses with the menu options of Biodiversity.R

Select the species and environmental matrices:

Biodiversity > Environmental Matrix > Select environmental matrix

→ Select the dune.env dataset

Biodiversity > Community Matrix > Select community matrix

→ Select the dune dataset

To calculate and plot a rank-abundance curve:

Biodiversity > Analysis of diversity > Rank abundance...

To model a rank-abundance curve:

Biodiversity > Analysis of diversity > Rank abundance...

→ Plot options: fit RAD

To calculate and plot a Rényi diversity profile:

Biodiversity > Analysis of diversity > Renyi profile...

→ Calculation method: all

To calculate and plot a Rényi diversity profile for each site separately:

Biodiversity > Analysis of diversity > Renyi profile...

→ Calculation method: separate per site

To calculate diversity indices for each site:

Biodiversity > Analysis of diversity > Diversity indices...

→ Diversity index: Shannon

→ Calculation method: separate per site

To compare diversity between subsets of the dataset:

Biodiversity > Analysis of diversity > Renyi profile...

→ Calculation method: separate per site

→ Subset options: Management

→ Subset: .

To calculate accumulation patterns for the Rényi diversity profile

Biodiversity > Analysis of diversity > Renyi profile...

→ Calculation method: accumulation

Doing the analyses with the command options of Biodiversity.R

To calculate and plot a rank-abundance curve:

```
RankAbun.1 <- rankabundance(dune)
RankAbun.1
rankabunplot(RankAbun.1, scale='abundance')
rankabunplot(RankAbun.1, scale='proportion')
```

To model a rank-abundance curve:

```
radfitresult(dune)
```

To calculate and plot a Rényi diversity profile:

```
Renyi.1 <- renyiresult(dune)
Renyi.1
renyiplot(Renyi.1)
renyiplot(Renyi.1, evenness=TRUE)
```

To calculate and plot a Rényi diversity profile for each site separately:

```
Renyi.2 <- renyiresult(dune, method='s')
Renyi.2
renyiplot(Renyi.2)
renyiplot(Renyi.2, evenness=TRUE)
```

To calculate diversity indices for each site:

```
Diversity.1 <- diversityresult(dune, index='Shannon',
,method='s')
Diversity.1
Diversity.2 <- diversityresult(dune, index='Simpson',
,method='s')
Diversity.2
Diversity.3 <- diversityresult(dune, index='Logalpha',
,method='s')
Diversity.3
```

To compare diversity between subsets of the dataset:

```
Renyi.3 <- renyicomp(dune, y=dune.env, factor='Management',
,permutations=100)
Renyi.3
```

To calculate accumulation patterns for the Rényi diversity profile

```
Renyi.4 <- renyiaccum(dune, permutations=100)
Renyi.4
renyiaccumplot(Renyi.4)
```