BORNH Bulletin of Regional Natural History

Formerly Bollettino della Societá dei Naturalisti in Napoli

Pareto model for species diversity

Stefano Giustino

DOI https://doi.org/10.6093/2724-4393/12712

***Correspondence:**
stefanogiustino@tiscali.it

Affiliation:
Via Posillipo 85, 80123 Napoli

Competing interests: The Author declares that they have no competing interests for this work.

Financial Disclosure Statement: the Author declares that no specific funding was received for this work.

Submitted: 08 May 2025
Revised: 05 Sept 2025
Accepted: 15 Sept 2025
Published: 26 Sept 2025

Associate Editor:
Marco Guida

Abstract

There are several mathematical models that describe the relationship between the number of species and the number of individuals per species. This article aimed to verify whether the Pareto mathematical model could be applied to ecological data regarding species and their abundance. The Pareto model has proven to be a valid descriptor of the relationship between species and the number of individuals per species, managing to measure the level of equality and/or dominance of the same. The Pareto model, therefore, is a candidate to be a valid aid for community ecologists in describing species diversity. In this paper, using a dataset taken from scientific literature I set out to verify the application of the Pareto model to the study of the relationship between species and abundance, with the ultimate aim of suggesting this method as an addition to those traditionally used.

Keywords: Species, Biodiversity, Pareto, Alpha, Models.

Riassunto

Esistono diversi modelli matematici che descrivono la relazione tra il numero di specie e il numero di individui per specie. Questo articolo mirava a verificare se il modello matematico di Pareto potesse essere applicato ai dati ecologici relativi alle specie e alla loro abbondanza. Il modello di Pareto si è dimostrato un valido descrittore della relazione tra le specie e il numero di individui per specie, riuscendo a misurare il livello di uguaglianza e/o dominanza delle stesse. Il modello di Pareto, quindi, è un valido aiuto per gli ecologi delle comunità nella descrizione della diversità delle specie. In questo articolo, utilizzando un set di dati tratto dalla letteratura scientifica, ho cercato di verificare l'applicazione del modello di Pareto allo studio della relazione tra

specie e abbondanza, con l'obiettivo finale di suggerire questo metodo come integrazione a quelli tradizionalmente utilizzati.

Parole chiave: Specie, Biodiversità, Pareto, Alfa, Modelli.

Introduction

In no natural community do all species have the same number of individuals, the observed pattern is always that of finding few species with many individuals and many species with few individuals (Magurran, 1988). This regularity in the relationship between species and their abundances has prompted the development of mathematical models that describe this relationship (Magurran, 1988). Mathematical models that relate the number of species to the number of individuals per species represent the most complete mathematical description of the data (Magurran, 1988). Currently, the mathematical models that describe the relationship between the number of species and the number of individuals for each species are mainly four: the Log-Normal series, the Geometric Series, the Logarithmic Series, the Broken Stick (Magurran, 1988; Krebs, 1999; Southwood & Henderson, 2000).
The Italian economist Vilfredo Pareto (1848-1923) devoted much time to the study of personal income inequality, in particular the observations suggested to Pareto the formulation of a universal law of income distribution, expressed in a hyperbolic mathematical model (Guarini & Tassinari, 1990). The Pareto distribution highlights how low incomes affect a greater number of people, compared to high incomes, which instead affect a minority of people. Furthermore, the parameter α present in the Pareto mathematical model, represents a real index of income concentration widely used in economics (Fraire & Rizzi, 2005).
If we consider the monetary income per person as equal to the number of individuals per species, and the income earners as the species, it is possible to adapt the Pareto model to ecological data regarding the relationship between the number of species and the number of individuals per species, furthermore the parameter α also allows us to quantify the level of concentration, or dominance, of the species in the community.
This work aims to apply the Pareto model for analyzing the relationship between the number of individuals per species and the number of species within Vertebrates using a data set reported in Magurran (1988). This work highlights that Pareto model can represent a useful analysis tool that integrates other mathematical models and indices such as Log-Normal series, the Geometric Series, the Logarithmic Series, the Broken Stick (Magurran, 1988; Krebs, 1999; Southwood & Henderson, 2000), so far used for studying the relationship between the number of individuals per species and the number of species.

Materials and methods

The Pareto mathematical model arises from the observation that the logarithm of the number of people is in inverse linear relation with the logarithm of income (Guarini & Tassinari, 1990); given N(x) the number of income earners and x the income, the relation is:

$$[1] \qquad \log N(x) = \log(K) - \alpha \log x$$

the ordinate at the origin is the log(K) while the slope is represented by the constant α. From [1] we obtain the Pareto law:

$$[2] \qquad N(x) = K / (x^\alpha)$$

defined for values of x included in the interval h and $+\infty$ with h corresponding to the initial level of the empirical distribution, with $h > 0$; $K > 0$ and $\alpha > 0$ (Guarini & Tassinari,

Table 1. Bird species abundance in remote European woodlands (from Magurran, 1988).

id	Species	Hidden Glen	Wild Wood	Lonely Pines
1	Spotted ratcatcher	1	2	0
2	Killer quail	3	16	354
3	Riff raff	2	3	7
4	Slyneck	1	2	4
5	Oat crake	4	10	29
6	Cold start	5	13	4
7	Big dipper	1	30	3
8	Shylark	1	14	12
9	Startling	18	22	18
10	Deadwing	1	1	2
11	Crook	2	4	1
12	Nightcap	63	5	1
13	Golden lover	2	19	1
14	Baby bunting	1	18	1
15	Mute swain	1	14	2
16	Chinese kite	1	15	0
17	Brownie owl	16	1	3
18	Hen hurrier	15	27	1
19	Grrr falcon	60	36	0
20	Gosh hawk	1	3	2
21	Cought	1	47	0
22	Flapping	8	38	18
23	Not	16	4	0
24	Bar-tailed nitwit	127	6	0
25	Snoop	9	7	0
26	Funny tern	18	8	1
27	Cut throat	3	16	0
28	Throttled dove	4	32	0
29	Ribbon	3	19	1
30	Backchat	11	6	1
31	Missile thrush	6	7	1
32	Cold tit	7	8	11
33	Twit	8	16	9
34	Yellow spanner	63	27	10
35	Born owl	17	4	3
	Σ =	500	500	500

1990). The value of h for the application that propose is equal to 1, a single individual per species. The constant α, the slope of the line, is used as an index of income concentration, with respect to income earners (Fraire & Rizzi, 2005). The parameter α is connected to the shape of the curve and is interpreted as a descriptive and direct index of the inequality of the distribution, in the sense that it increases as the inequality (concentration) of income increases (Guarini & Tassinari, 1990). The minimum value of the slope is 0 in this limiting case N(x) = K, which represents perfect equal distribution of individuals across species. The maximum value of the slope, on the other hand, is undefined, because the tangent of an angle grows to infinity. Table 1 reports the data on which the Pareto model was applied, taken from Magurran (1988). Thirty-five species of birds are represented with their respective abundances resulting from reserache carried out in three differnt locations. For each series of data referring to a location, a distribution of the number of species for the equally large classes of individuals was first constructed. Once the distribution of species-abundance classes was obtained, for each class its central value was considered, so as to obtain an ordered series of points in a Cartesian plane. This operation was necessary in order to estimate the parameters of the model [1], in fact with these series of points by applying the ordinary least squares (OLS) estimation method regressing the log of the number of species on the log of the abundances of individual species, it was possible to obtain the value of Log(K) and α for each of the three locations. The logarithms used are the Neperian ones. Once Log(K) was obtained, it was possible to easily obtain K = Exp[Log(K)], while for α the negative sign was not considered. With the obtained parameters K and α, the model [2] was applied to the real data, always using the central value of the class for the variable x, in order to obtain the theoretical values to be compared with the observed values. Finally, to evaluate the goodness of the adaptation of the real data to the theoretical data predicted by the model, the non-parametric adaptation test for a theoretical distribution of Kolgomorov-Smirnov was used (Siegel & Castellan, 1988).

Results

Location of hidden glen is given in Fig. 1; the value of K = 33,87 and α = 0,77 , the value of the fit statistic D = 0,041 n.s. the theoretical model fits the real data. Location of wild wood is given in Fig. 2; the value of K = 29,13 and

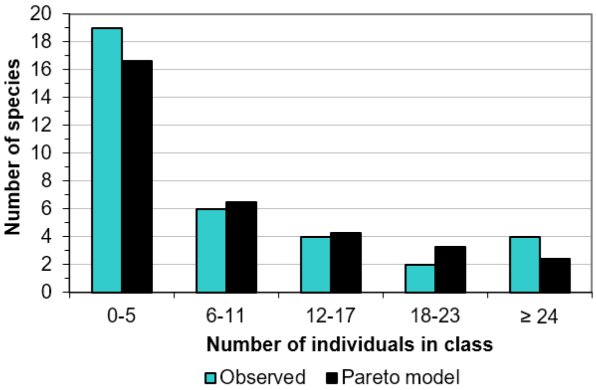

Figure 1. Hidden glenn, observed values and theoretical values of the number of species per abundance class.

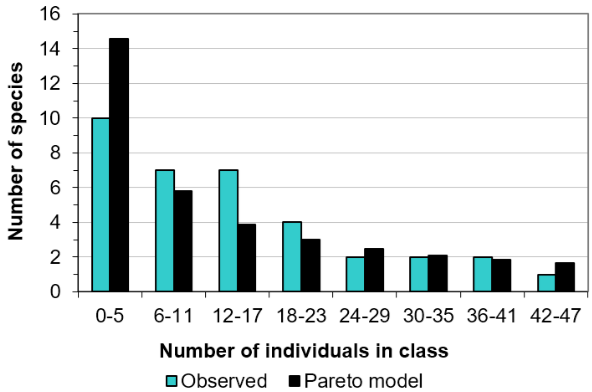

Figure 2. Wild wood, observed values and theoretical values of the number of species per abundance class.

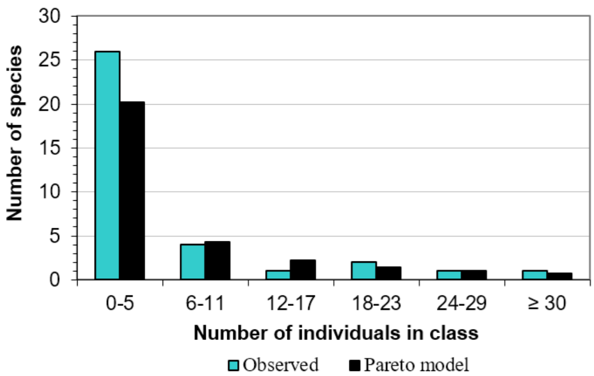

Figure 3. Lonely Pines, observed values and theoretical values of the number of species per abundance class.

α = 0,75 , the value of the fit statistic D = 0,13 n.s. the theoretical model fits the real data. Location of lonely pines is given in Fig. 3; the value of K = 63,78 and α = 1,26 , the value of the fit statistic D = 0,069 n.s. the theoretical model fits the real data. The value of α for the Lonely Pines locality is higher than the value of α for the other two localities, and this is also clearly highlighted by the graph, therefore the level of concertation of individuals per species in the Lonely Pines locality is higher than that of the other localities. The two locations Wild Wood and Hidden Glen have an equal value of α and consequently an equal diversity of species. The value for K and α all three examined localities is reported in Table 2.

Table 2. Parameters estimated using. Pareto model: For details see text

Pareto model	Hidden Glen	Wild Wood	Lonely Pines
K	33,9	29,1	63,8
α	0,77	0,75	1,26

Discussion

To the best of my knowledge, this work represents one of the first study that applies the Pareto Model for analyzing the relationships between the number of species and the number of individuals per species in vertebrates. In fact, this relationship is generally addressed using many other models such as those previously mentioned (Magurran, 1988; Krebs, 1999; Southwood & Henderson, 2000), or more recent models obtained using ad hoc algorithms (Yingnan et al. 2025; Chave, 2024; Moradi et al. 2025).
In general, species diversity measures take into account two quantities: the number of species and the number of individuals per species, the latter also known as evenness or equitability, furthermore the greater the regularity of the number of individuals for each species, the greater the diversity of a commu-

nity (Magurran, 1988). Mathematical models that describe the relationship between the number of species and the number of individuals per species are able to highlight when there is a situation of concentration or equidistribution of individuals in species, and furthermore if a single diversity index is needed, a parameter of an appropriate species-abundance distribution can be used (Magurran, 1988). The Pareto model, born to measure and describe the concentration/inequality of income for human societies (Guarini & Tassinari, 1990), has adapted well to data regarding the biodiversity of bird species, and this is because the principle on which the Pareto model is based is universal. Furthermore, the parameter α of the Pareto model, which is a direct descriptive index of the inequality of distribution (Guarini & Tassinari, 1990), has highlighted how among the three locations examined there is a difference in the diversity of species with a better evenness in the locations of Wild Wood and Hidden and a worse diversity in the location of Lonely Pines. A greater concentration of individuals in a few species means less diversity and greater dominance of a few species, when one or a few species are prevalent within the community, these organisms are therefore defined (taxa) dominant; dominance is the opposite of diversity (Smith & Smith, 2006).

References

Chave J. (2024). *Species abundance, urn models, and neutrality* Academie des sciences Istitut de France Volume 347 (2024), p. 119-135 https://doi.org/10.5802/crbiol.162

Fraire M., Rizzi A. (2005). *Statistica: metodi esplorativi ed inferenziali* Carocci Editore.

Guarini R., Tassinari F. (1990). *Statistica economica.* Il Mulino, Bologna.

Krebs J.C. (1999). *Ecological Methodology* Addison Wesley Longman, Inc.

Magurran A. E. (1988). *Ecological diversity and its measurement* Princeton University Press.

Moradi H., Inácio P.P., Oldeland J., Borges P.A.V., Fattorini S. (2025). Use of cover data to model species abundance distributions through continuous probability functions *Scientific Reports*, **15**:11361 | https://doi.org/10.1038/s41598-025-94587-w

Siegel S., Castellan N.J. (1988). *Nonparametrics Statistics for the Behavioral Sciences* Mc-Graww-Hill.

Southwood T.R.E., Henderson P.A. (2000). *Ecological Methods* 3d Edition Blackwell Science, XV+5775 pp.

Smith T.M., Smith R.L. (2006). *Elements of ecology* Pearson Education, Inc.

Yingnan G. Ahmed A., MartinW. (2025). The powerbend distribution provides a unified model for the species abundance distribution across animals, plants and microbes. *Nature Communication* Volume 16, p. 4035 https://doi.org/10.1038/s41467-025-59253-9

Bulletin of Regional Natural History (BORNH) ISSN 2724-4393.